The copositive cone, the completely positive cone and their generalisations

Dickinson, Peter James Clair

The Copositive Cone, the Completely Positive Cone and their Generalisations

Peter J.C. Dickinson

rijksuniversiteit
groningen

This PhD project was carried out at the Johann Bernoulli Institute according to the requirements of the Graduate School of Science (Faculty of Mathematics and Natural Sciences, University of Groningen).

Funding was provided to me as a bursary student from the Rosalind Franklin Fellowship programme of the University of Groningen.

Rijksuniversiteit Groningen

The Copositive Cone, the Completely Positive Cone and their Generalisations

Proefschrift

ter verkrijging van het doctoraat in de
Wiskunde en Natuurwetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. E. Sterken,
in het openbaar te verdedigen op
vrijdag 12 april 2013
om 16.15 uur

door

Peter James Clair Dickinson

geboren op 16 maart 1985
te Ealing, Engeland

Promotor:	Prof. dr. M. Dür
Beoordelingscommissie:	Prof. dr. I.M. Bomze
	Prof. dr. M. Laurent
	Prof. dr. A.N. Letchford

ISBN: 978-90-367-6101-7 (printed version)
ISBN: 978-90-367-6107-9 (electronic version)

Contents

Acknowledgements **v**

Personal contributions **vii**

I Introduction **1**

1 Introduction **3**

1.1 Copositive & Completely Positive Cones 3
1.2 Geometry and Optimisation 7
1.2.1 Geometry 7
1.2.2 Mathematical Optimisation 10
1.2.3 Duality 12
1.2.4 Special types of conic optimisation 17
1.3 Maximum Weight Clique Problem 18

2 Set-semidefinite Optimisation **23**

2.1 Set-semidefinite Cone 23
2.2 Single Quadratic Constraint Problem 24
2.3 Standard Quadratic Optimisation Problem 26
2.4 Quadratic Binary Optimisation 27
2.5 Considering a special case 32

3 Complexity **35**

3.1 Membership Problems 35
3.2 Ellipsoid Method 37
3.3 Copositivity and Completely Positivity 41

4 Sparse $\mathcal{C}^*$ detection and decomposition **49**

4.1 Rank-One Decomposition 51
4.2 Indices of Degree Zero or One 53
4.3 Chains 57

4.4 Matrices with circular graphs 60
4.5 Reducing Chain Lengths 64
4.6 Preprocessing 68
4.7 Number of minimal decompositions 69

II Geometry **73**

5 Proper Cones **75**

6 The Set of Zeros **77**

7 Interior of $\mathcal{C}$ and $\mathcal{C}^*$ **79**
7.1 Introduction 79
7.2 Original Proof 81
7.3 New Proof 82
7.4 Concluding Remarks 83

8 Facial Structure **85**
8.1 Geometry of General Proper Cones 85
8.2 $\mathcal{C}^2$ and $\mathcal{C}^{*2}$ 89
8.3 Extreme Rays of $\mathcal{C}$ and $\mathcal{C}^*$ 90
8.4 Maximal Faces of the Copositive Cone 95
8.5 Maximal Faces of the Completely Positive Cone 98
8.6 Lower Bound on Dimension of Maximal Faces of $\mathcal{C}^*$ 100

III Approximations **103**

9 Simplicial Partitioning **107**
9.1 Introduction 107
9.1.1 Approximation hierarchy 108
9.1.2 Partitioning methods 110
9.2 Counter-examples 112
9.3 An Exhaustive Partitioning Scheme 118
9.4 Reconsidering unrestricted free bisection 120
9.5 The importance of picking a fixed lambda or rho 121
9.6 Conclusion 124

10 Moment Approximations **125**
10.1 Introduction 125
10.1.1 Notation 125
10.1.2 Contribution 126

10.2 Introduction to Moments 126
10.3 Lasserre's hierarchies 127
10.4 New hierarcies 129
10.5 Examples 131
10.6 Conclusion 133

11 Sum-of-Squares **135**
11.1 Introduction 135
11.2 Parrilo's approximation hierarchy 137
11.3 Scaling a matrix out of $\mathcal{K}_n^r$ 137
11.4 Scaling a matrix into $\mathcal{K}_5^1$ 138
11.5 The importance of scaling to binary diagonals 140
11.6 Dual of the Parrilo approximations 141

12 Cones of Polynomials **145**
12.1 Positivstellensätze 145
12.2 Reconsidering polynomial constraints 147
12.3 Application to Optimisation 147
12.4 Inner Approximation Hierarchy 149

Summary **151**

Samenvatting **153**

Bibliography **155**

Nomenclature **163**

Index **167**

Acknowledgements

I have found my time as a PhD student to be both enjoyable and profitable. However, this would not have been possible without the help that I received from a great many people, and I would like to take this opportunity to thank some of them.

First and foremost, I would like to thank my supervisor, Prof. Dr. Mirjam Dür. When coming to Mirjam to ask for the PhD position, I had never studied optimisation before and had been away from academia for two years. However, she must have seen some potential in me and gave me a chance, something that I will be eternally grateful for. Naturally, I would like to thank her for teaching me the subject matter that I have needed to know, however she has also done much more than that. When I started to write up my work, I produced little more than a mess of equations, but Mirjam taught me to write in a more legible manner. She also taught me how academia works and provided lots of advice for my future career. Most importantly of all, she made me feel welcome, and made my time as a PhD student an enjoyable one.

There are many people that I would like to thank at the University of Groningen and in particular at the Johann Bernoulli Institute. I would like to thank all of my colleagues, who are too numerous to include a list of. Our lunches always gave me a nice break during the day. One person that I would like to include extra thanks to is Luuk Gijben for our enjoyable collaborations together and for translating the summary of this thesis into Dutch.

In addition to my colleagues in Groningen, I would also like to thank my colleagues from around the world. Without them, the conferences that I went to would not have been as informative or enjoyable as they were. I wish to especially thank my reading committee, Prof. Dr. Immanuel Bomze, Prof. Dr. Monique Laurent and Prof. Dr. Adam Letchford, for the time and care that they took in considering my thesis. I would further like to thank colleagues that I have collaborated with during my PhD. Many of these colleagues I also visited, during which they made me feel very welcome. These colleagues are: Prof. Dr. Kurt Anstreicher, Prof. Dr. Samuel Burer, Prof. Dr. Gabriele Eichfelder, Dr. Roland Hildebrand and Prof. Dr. Janez Povh.

Organisations that I would like to thank for making this PhD possible are the University of Groningen, the Johann Bernoulli Institute, the Groningen Graduate School of Science and the Landelijk Netwerk Mathematische Besliskunde. I am also thankful for the funding that I received as a bursary student from the Rosalind Franklin Fellowship programme of the University of Groningen. Although I am grateful to these organisation, most people who know me will know that I have something verging on a hatred for forms and other aspects of dealing with administration. I would thus like to thank those who have helped to reduce my administrative burden, especially Esmee Elshof and Ineke Schelhaas. In addition to this, I would also like to thank Peter Arendz for his regular assistance in all things connected to computers.

There also many people from outside of academia who have provided help to me. I am grateful for the support shown to me throughout my life by my family, especially Keith Dickinson, Claire Dickinson and Katie Dickinson. I would also like to thank the family of my girlfriend, especially Erik Bosman, Geertje Bosman-Wichers and Marije Bosman, who provided invaluable support from the start of my life in the Netherlands. I would also like to thank my friends in both England and the Netherlands for helping me to relax. In the Netherlands, this was primarily through the scuba diving club G.B.D. Calamari, who showed me friendship as well as providing me a place to continue my hobby of scuba diving.

One final very big thanks goes to my girlfriend, Renske Bosman. Without her support through both the good and the tough times of my PhD, I never would have managed. I hope that I can show the same support to her when she does a PhD.

Personal contributions

I began my PhD with Prof. Dr. Mirjam Dür at the University of Groningen in April 2009, and during my time as a PhD student I have produced (along with my co-authors) the following articles:

Published articles:

[Dic13] Peter J.C. Dickinson. On the Exhaustivity of Simplicial Partitioning. *Journal of Global Optimization*, in print. DOI: 10.1007/s10898-013-0040-7

[DP13a] Peter J.C. Dickinson and Janez Povh. Moment approximations for set-semidefinite polynomials. *Journal of Optimization Theory and Applications.* DOI: 10.1007/s10957-013-0279-7

[DDGH13] Peter J.C. Dickinson, Mirjam Dür, Luuk Gijben and Roland Hildebrand. Scaling relationship between the copositive cone and Parrilo's first level approximation. *Optimization Letters*, in print. DOI: 10.1007/s11590-012-0523-3.

[DD12] Peter J.C. Dickinson and Mirjam Dür. Linear-time complete positivity detection and decomposition of sparse matrices. *SIAM Journal on Matrix Analysis and Applications*, 33(3):701–720, 2012.

[Dic11] Peter J.C. Dickinson. Geometry of the copositive and completely positive cones. *Journal of Mathematical Analysis and Applications*, 380(1):377–395, 2011.

[Dic10] Peter J.C. Dickinson. An improved characterisation of the interior of the completely positive cone. *Electronic Journal of Linear Algebra*, 20:723–729, 2010.

Accepted articles:

[DDGH12] Peter J.C. Dickinson, Mirjam Dür, Luuk Gijben and Roland Hildebrand. Irreducible elements of the copositive cone. *Linear Algebra and its Applications.*

[DEP12] Peter J.C. Dickinson, Gabriele Eichfelder and Janez Povh. Erratum to “On the set-semidefinite representation of nonconvex quadratic programs over arbitrary feasible sets” [Optim. Letters, 2012]. *Optimization Letters.*

Conditionally accepted, minor revisions:

[DG11] Peter J.C. Dickinson and Luuk Gijben. On the computational complexity of membership problems for the completely positive cone and its dual. *Computational Optimization and Applications.*

Submitted articles:

[DP13c] Peter J.C. Dickinson and Janez Povh. On a generalization of Pólya’s and Putinar-Vasilescu’s Positivstellensätze. *Journal of Global Optimization.*

Articles in construction:

[ABD12] Kurt M. Anstreicher, Samuel Burer and Peter J.C. Dickinson. An algorithm for computing the CP-factorization of a completely positive matrix.

[DP13b] Peter J.C. Dickinson and Janez Povh. A new convex reformulation and approximation hierarchy for polynomial optimization.

Rather than simply stapling these papers together to provide a thesis, I have instead tried to provide a self-contained text on the field of copositive optimisation. This necessarily involves including results which are not part of my own work. I will thus use this chapter to explain my personal contributions.

Chapter 1 acts as a general introduction to copositive optimisation. This is primarily a review of well-known results. The only exception is Section 1.3, which provides a new proof for the copositive reformulation of the maximum weight clique problem.

Chapter 2 expands on my paper [DEP12], which was written with Prof. Dr. Janez Povh and Prof. Dr. Gabriele Eichfelder. This paper was correcting an error that I found in the paper [EP12]. For this reason, we primarily focused on the results from the original paper and used a similar structure of proof. In this thesis, I have decided to look at a new proof of this, which I find more natural and allows for more applications.

Chapter 3 looks at the complexity results connected to the copositive and completely positive cones from my papers [DG11] and [ABD12], which I wrote with Luuk Gijben, Prof. Dr. Kurt Anstreicher and Prof. Dr. Samuel Burer. In these papers, numerous results were shown which were largely expected to be true, however the technical details had not previously been analysed. The first two sections in this chapter act as an introduction to some standard definitions, methods and results. Then, in the final section, rather than extensively going through all of the technical results from the papers, only those required from [ABD12] are looked at. This should then give the reader a flavour of the techniques that were involved.

Chapter 4 is the final chapter of Part I, and this looks at the results from my paper [DD12], which I wrote with Prof. Dr. Mirjam Dür. This paper showed that even though checking whether a matrix is completely positive is an $\mathcal{NP}$-hard problem, we are able to check certain special types of sparse matrices in linear time. Related work had previously been done in the past, however our work provided a unified approach to these ideas, extended the domain of cases that can be considered and analysed the running-time. A thorough discussion of previous work is provided near the start of this chapter.

In Part II, we look at geometric properties of the copositive and completely positive cones. Chapter 5 provides a proof of the well-known result that both the copositive and completely positive cones are proper cones. A proof is provided due to the importance of this result and as a simple problem to get the reader thinking about geometry.

Chapter 6 looks at the set of zeros. Although this set is not directly connected to geometry, it has come in very useful when considering geometric properties. This idea originates from [Dia62], and was further developed by myself in the articles [Dic10] and [Dic11].

Chapter 7 looks at the results from my paper [Dic10] on the interior of the completely positive cone, as well as providing a new proof for these results.

Chapter 8 is the final chapter of Part II, and this considers results on the facial structure of the copositive and completely positive cones from my paper [Dic11]. This builds on previous work on extreme rays of the copositive and completely positive cones from various articles.

Due to the complexity of copositive optimisation, it is preferable to consider optimisation over approximations of the copositive cone, rather than trying to optimise over the cone directly. These approximations are considered in Part III, starting with a general introduction on well-known methods and results.

After this general introduction, we look at four types of approximations of the copositive cone, starting in Chapter 9 with simplex partitioning. This provides two methods in connection to copositivity, which were introduced in the papers [BD08] and [BE12]. However, in my paper [Dic13], it was shown that results stated in these papers on the convergence of their algorithms are incorrect, and it is the analysis from [Dic13] that will be considered in this chapter. Numerous cases of when we would expect convergence to occur are considered, and counterexamples are constructed to show that we do not get this expected convergence. A new method for simplex partitioning is also introduced, and from analysis of this we are able to give some fairly relaxed conditions for convergence.

In Chapter 10 we look at the results from my paper [DP13a], which I wrote with Prof. Dr. Janez Povh. This analyses an approximation method connected to moments, which was introduced in [Las11]. We provided a new way of considering this method which not only provides new insights, but also allows us to introduce a new approximation method which is at least as good as that provided by Lasserre.

Closely connected to the idea of moments is that of sums-of-squares, which is looked at in Chapter 11. We start by using standard analysis to consider how, through duality, the method from [Las11] is connected to sums-of-squares. This means that in effect Lasserre's method is using sums-of-squares. However, a more common way of using sums-of-squares is that introduced in [Par00] in the form of Parrilo-cones, and Section 11.2 briefly reviews this method. Sections 11.3 to 11.5 consider results from my paper [DDGH13], which I wrote with Prof. Dr. Mirjam Dür, Luuk Gijben and Dr. Roland Hildebrand. This includes surprising results on the Parrilo-cones, with Section 11.4 building on results from our paper [DDGH12]. In the final section of this chapter, Section 11.6, standard analysis is used to consider the dual of the Parrilo-cones, which shows how it is connected to moments. Although, as far as I am aware, most of the results on the duals from Sections 11.1 and 11.6 have not previously appeared in the literature, I do not feel that I can honestly claim them to be new results.

Connected to the Parrilo-cones are two alternative methods for approximating the copositive cone which were introduced in [BK02] and [PVZ07]. In my papers [DP13b, DP13c], which I wrote with Prof. Dr. Janez Povh, we show that by constructing a new positivstellensatz, we are able to provide a new approximation method for a generalisation of the copositive cone, which includes these two previous approximation methods as special cases. In Chapter 12, results on this new positivstellensatz and the associated approximation scheme are considered.

Part I

Introduction

Chapter 1

Introduction

The majority of the notation used in this thesis will be standard notation. Rather than define this within the text, the reader is directed towards the nomenclature provided at the end of this thesis. A bibliography and index are also provided at the end of this thesis.

In this chapter, we introduce the copositive and completely positive cones, as well as their relation to conic optimisation.

1.1 Copositive & Completely Positive Cones

A symmetric matrix A is defined to be *copositive* if $\mathbf{x}^\mathsf{T} A\mathbf{x} \geq 0$ for all nonnegative vectors $\mathbf{x}$, and we denote the *copositive cone* by

$$\mathcal{C}^n := \{A \in \mathcal{S}^n \mid \mathbf{x}^\mathsf{T} A\mathbf{x} \geq 0 \text{ for all } \mathbf{x} \in \mathbb{R}^n_+\}.$$

A symmetric matrix A is defined to be *completely positive* if there exists a nonnegative matrix B such that $A = BB^\mathsf{T}$, and we denote the *completely positive cone* by

$$\begin{aligned}\mathcal{C}^{*n} &:= \{BB^\mathsf{T} \in \mathcal{S}^n \mid B \text{ is a nonnegative matrix}\}\\ &= \{\textstyle\sum_i \mathbf{a}_i\mathbf{a}_i^\mathsf{T} \mid \mathbf{a}_i \in \mathbb{R}^n_+ \text{ for all } i\}.\end{aligned}$$

We define the dual of a set $\mathcal{K} \subseteq \mathcal{S}^n$ by

$$\mathcal{K}^* := \{X \in \mathcal{S}^n \mid \operatorname{trace}(AB) \geq 0 \text{ for all } Y \in \mathcal{K}\}.$$

Using this definition, we get that the copositive and completely positive cones are mutually dual to each other, and it is this property that motivates their combined study.

In spite of this connection, the concepts of copositivity and complete positivity were in fact introduced in different areas of mathematics. The concept of copositivity was first introduced in the field of linear algebra by Prof. Dr. Theodore S. Motzkin in the 1950s [Mot52], whilst the concept of complete positivity was first introduced in the field of numerical/combinatorial analysis by Prof. Dr. Marshall Hall Jr. in the 1960s [HJ62]. However, Hall did note the connection between the copositive and completely positive cones.

In this thesis, we shall be looking at these cones from the viewpoint of mathematical optimisation. This is due to the important applications of these cones in optimisation which were discovered during the last 15 years [QKRT98, BDK$^+$00, KP02, Bur09].

In the following section, we shall look at basic properties of geometry and optimisation, which will give us the tools to look at one of the major motivations of the study of optimisation in connection with the copositive and completely positive cones, which we shall do in the final section of this chapter.

In Chapter 2, we shall look at a generalisation of these cones, focusing on applications of this generalisation. Then, in Chapter 3, we will review results on the complexity of checking whether a matrix is copositive, and look at similar results on the complexity of checking whether a matrix is completely positive. This is in fact an $\mathcal{NP}$-hard problem, however, in Chapter 4, we shall show that we are able to check whether certain types of sparse matrices are completely positive in linear time.

In Part II, we shall look at geometric properties of these cones in order to improving our understanding of them, whilst, in Part III, we shall look at approximations to these cones (and their generalisations). A bibliography, nomenclature and index are then provided at the end of this thesis.

However, before we continue with the rest of this thesis, we shall first look at some basic properties of copositive and completely positive matrices which will be required throughout.

We define the set of nonnegative symmetric matrices and the *positive semidefinite cone* respectively as follows:

$$\begin{aligned}\mathcal{N}^n &:= \{A \in \mathcal{S}^n \mid (A)_{ij} \geq 0 \text{ for all } i, j\},\\ \mathcal{S}^n_+ &:= \{A \in \mathcal{S}^n \mid \mathbf{x}^\mathsf{T} A \mathbf{x} \geq 0 \text{ for all } \mathbf{x} \in \mathbb{R}^n\}\\ &= \{\textstyle\sum_i \mathbf{a}_i \mathbf{a}_i^\mathsf{T} \mid \mathbf{a}_i \in \mathbb{R}^n \text{ for all } i\}.\end{aligned}$$

From the definitions, it is immediately apparent that $\mathcal{C}^{*n} \subseteq \mathcal{S}^n_+ \cap \mathcal{N}^n$ and that $\mathcal{S}^n_+ + \mathcal{N}^n \subseteq \mathcal{C}^n$. It has in fact been shown in [MM62] that these inclusions hold with equality for $n \leq 4$ and are strictly for $n \geq 5$. We refer to $\mathcal{S}_+ \cap \mathcal{N}$ as the doubly nonnegative cone.

In the following two theorems we look at further basic properties of the copositive and completely positive cones.

Theorem 1.1. *Let A be a copositive matrix. Then we have that:*

i. *If P is a permutation matrix and D is a nonnegative diagonal matrix, then $PDADP^\mathsf{T} \in \mathcal{C}$.*

ii. *Every principal submatrix of A is also copositive (where a principal submatrix of A is a matrix formed by deleting any rows along with the corresponding columns from A).*

iii. *$(A)_{ii} \geq 0$ for all i.*

iv. *If $(A)_{ii} = 0$, then $(A)_{ij} \geq 0$ for all j.*

v. *$(A)_{ij} \geq -\sqrt{(A)_{ii}(A)_{jj}}$ for all i, j.*

vi. *If there exists a strictly positive vector $\mathbf{v}$ such that $\mathbf{v}^\mathsf{T} A\mathbf{v} = 0$, then $A \in \mathcal{S}_+$.*

Proof. (i) is a well-known result and proofs of (ii)–(vi) are provided in [Bun09, Dia62]. However, for the sake of completeness, we shall give brief proofs of these results here. For this we consider an arbitrary matrix $A \in \mathcal{C}^n$.

i. We have that $DP^\mathsf{T}\mathbb{R}^n_+ \subseteq \mathbb{R}^n_+$, and thus for all $\mathbf{x} \in \mathbb{R}^n_+$ we have

$$0 \leq (DP^\mathsf{T}\mathbf{x})^\mathsf{T} A(DP^\mathsf{T}\mathbf{x}) = \mathbf{x}^\mathsf{T}(PDADP^\mathsf{T})\mathbf{x}.$$

ii. Let $A_1 \in \mathcal{S}^m$ be a principal submatrix of A. From (i), without loss of generality, we may assume that $A = \begin{pmatrix} A_1 & A_2^\mathsf{T} \\ A_2 & A_3 \end{pmatrix}$, where $A_2 \in \mathbb{R}^{(n-m)\times m}$ and $A_3 \in \mathcal{S}^{(n-m)}$. We now note that for all $\mathbf{v} \in \mathbb{R}^m_+$ we have

$$0 \leq \begin{pmatrix} \mathbf{v} \\ \mathbf{0} \end{pmatrix}^\mathsf{T} \begin{pmatrix} A_1 & A_2^\mathsf{T} \\ A_2 & A_3 \end{pmatrix} \begin{pmatrix} \mathbf{v} \\ \mathbf{0} \end{pmatrix} = \mathbf{v}^\mathsf{T} A_1 \mathbf{v}.$$

iii. For all $i = 1, \ldots, n$ we have $\boldsymbol{e}_i \in \mathbb{R}^n_+$ and thus $0 \leq \boldsymbol{e}_i^\mathsf{T} A\boldsymbol{e}_i = (A)_{ii}$.

iv. Suppose for the sake of contradiction that there exists $i, j \in \{1, \ldots, n\}$ such that $(A)_{ii} = 0$ and $(A)_{ij} < 0$. From (iii) we have $(A)_{jj} \geq 0$ and thus $\big(((A)_{jj} + 1)\boldsymbol{e}_i - (A)_{ij}\boldsymbol{e}_j\big) \in \mathbb{R}^n_+$. This then gives the contradiction

$$\begin{aligned} 0 &\leq \big(((A)_{jj} + 1)\boldsymbol{e}_i - (A)_{ij}\boldsymbol{e}_j\big)^\mathsf{T} A\big(((A)_{jj} + 1)\boldsymbol{e}_i - (A)_{ij}\boldsymbol{e}_j\big) \\ &= -((A)_{jj} + 2)(A)_{ij}^2 < 0. \end{aligned}$$

v. From (*iii*) we get that $\big((A)_{ii}(A)_{jj}\big) \in \mathbb{R}_+$ for all $i, j = 1, \ldots, n$, and thus the inequality is well defined. Now suppose for the sake of contradiction that there exists $i, j \in \{1, \ldots, n\}$ such that $(A)_{ij} < -\sqrt{(A)_{ii}(A)_{jj}}$. This is equivalent to having $(A)_{ij} < 0$ and $(A)_{ij}^2 > (A)_{ii}(A)_{jj}$. From (*iii*),(*iv*) we get $(A)_{ii}, (A)_{jj} > 0$. Therefore $\big((A)_{jj}\boldsymbol{e}_i - (A)_{ij}\boldsymbol{e}_j\big) \in \mathbb{R}^n_+$. We now note the following contradiction:

$$\begin{aligned} 0 &\leq \big((A)_{jj}\boldsymbol{e}_i - (A)_{ij}\boldsymbol{e}_j\big)^\mathsf{T} A\big((A)_{jj}\boldsymbol{e}_i - (A)_{ij}\boldsymbol{e}_j\big) \\ &= (A)_{jj}\big((A)_{ii}(A)_{jj} - (A)_{ij}^2\big) < 0. \end{aligned}$$

vi. Suppose there exists $\mathbf{v} \in \mathbb{R}^n_{++}$ such that $\mathbf{v}^\mathsf{T} A\mathbf{v} = 0$, and consider an arbitrary $\mathbf{u} \in \mathbb{R}^n$. There exists $\widehat{\varepsilon} > 0$ such that $(\mathbf{v} \pm \varepsilon\mathbf{u}) \in \mathbb{R}^n_+$ for all $\varepsilon \in (0, \widehat{\varepsilon}]$. Therefore, for all $\varepsilon \in (0, \widehat{\varepsilon}]$ we have

$$\begin{aligned} 0 &\leq \tfrac{1}{\varepsilon}(\mathbf{v} + \varepsilon\mathbf{u})^\mathsf{T} A(\mathbf{v} + \varepsilon\mathbf{u}) = 2\mathbf{u}^\mathsf{T} A\mathbf{v} + \varepsilon\mathbf{u}^\mathsf{T} A\mathbf{u}, \\ 0 &\leq \tfrac{1}{\varepsilon}(\mathbf{v} - \varepsilon\mathbf{u})^\mathsf{T} A(\mathbf{v} - \varepsilon\mathbf{u}) = -2\mathbf{u}^\mathsf{T} A\mathbf{v} + \varepsilon\mathbf{u}^\mathsf{T} A\mathbf{u}. \end{aligned}$$

Considering $\varepsilon \to 0$, we get $\mathbf{u}^\mathsf{T} A\mathbf{v} = 0$, and thus $0 \leq \varepsilon\mathbf{u}^\mathsf{T} A\mathbf{u}$, or equivalently $0 \leq \mathbf{u}^\mathsf{T} A\mathbf{u}$. As $\mathbf{u} \in \mathbb{R}^n$ was arbitrary, this implies $A \in \mathcal{S}_+$, completing the proof. □

Theorem 1.2. *Let A be a completely positive matrix. Then we have that:*

i. If P is a permutation matrix and D is a nonnegative diagonal matrix, then $PDADP^\mathsf{T} \in \mathcal{C}^*$.

ii. Every principal submatrix of A is also completely positive.

iii. If $(A)_{ii} = 0$, then $(A)_{ij} = 0$ for all j.

Proof. Although these results are somewhat trivial, for the sake of completeness, we shall again give brief proofs. For this we consider arbitrary matrices $A \in \mathcal{C}^{*n}$ and $B \in \mathbb{R}^{n\times p}_+$ such that $A = BB^\mathsf{T}$.

i. We have $(PDB) \in \mathbb{R}^{n\times p}_+$ and thus $PDADP^\mathsf{T} = (PDB)(PDB)^\mathsf{T} \in \mathcal{C}^*$.

ii. Let $A_1 \in \mathcal{S}^m$ be a principal submatrix of A. From (*i*), without loss of generality, we may assume that $A = \begin{pmatrix} A_1 & A_2^\mathsf{T} \\ A_2 & A_3 \end{pmatrix}$, where $A_2 \in \mathbb{R}^{(n-m)\times m}$ and $A_3 \in \mathcal{S}^{(n-m)}$. We now let $B_1 \in \mathbb{R}^{m\times p}_+$ and $B_2 \in \mathbb{R}^{(n-m)\times p}_+$ such that $B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}$. We then have $A_1 = B_1B_1^\mathsf{T} \in \mathcal{C}^*$.

iii. We have $0 = (A)_{ii} = \sum_{k=1}^p (B)_{ik}^2$, and thus $(B)_{ik} = 0$ for all k. Therefore, for all j we have $(A)_{ij} = \sum_{k=1}^p (B)_{ik}(B)_{jk} = 0$. □

1.2 Geometry and Optimisation

This section will act as a brief introduction to some basic results on geometry and (conic) optimisation, which will be needed throughout this thesis. For a broader understanding, the books [Roc70, BV04] are recommended. We shall be considering Euclidean spaces, $\mathbb{R}^n$, with inner product defined as $\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{x}^\mathsf{T}\mathbf{y} = \sum_i (\mathbf{x})_i (\mathbf{y})_j$ and 2-norm defined as $\|\mathbf{x}\|_2 := \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. However, the definitions and results also trivially extend to the space of symmetric matrices, $\mathcal{S}^n$, with $\langle X, Y \rangle := \operatorname{trace}(XY) = \sum_{i,j} (X)_{ij} (Y)_{ij}$ and $\|X\|_2 := \sqrt{\langle X, X \rangle}$ for all $X, Y \in \mathcal{S}^n$.

Before doing this, we first consider the following well-known results on inner products and 2-norms.

Theorem 1.3. *For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $A, B \in \mathcal{S}^n$ we have*

$$\|\mathbf{x} + \mathbf{y}\|_2 \leq \|\mathbf{x}\|_2 + \|\mathbf{y}\|_2, \tag{1.1}$$
$$\|A + B\|_2 \leq \|A\|_2 + \|B\|_2, \tag{1.2}$$
$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\|_2 \|\mathbf{y}\|_2, \tag{1.3}$$
$$|\langle A, B \rangle| \leq \|A\|_2 \|B\|_2, \tag{1.4}$$
$$\|A\mathbf{x}\|_2 \leq \|A\|_2 \|\mathbf{x}\|_2, \tag{1.5}$$
$$|\mathbf{x}^\mathsf{T} A \mathbf{y}| \leq \|A\|_2 \|\mathbf{x}\|_2 \|\mathbf{y}\|_2. \tag{1.6}$$

Proof. (1.1) and (1.2) are simply the *Minkowski inequality*, whilst (1.3) and (1.4) are simply the *Cauchy-Schwarz inequality*, see for example [Ber09, Lemma 9.1.3 and Corollary 9.1.7].

Letting $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^n$ be the rows of A, we then have the following, which proves (1.5):

$$\|A\mathbf{x}\|_2^2 = \textstyle\sum_i \langle \mathbf{a}_i, \mathbf{x} \rangle^2 \leq \sum_i \|\mathbf{a}_i\|_2^2 \|\mathbf{x}\|_2^2 = \|A\|_2^2 \|\mathbf{x}\|_2^2.$$

Finally, we have the following, which proves (1.6):

$$|\mathbf{x}^\mathsf{T} A \mathbf{y}| = |\langle \mathbf{x}, A\mathbf{y} \rangle| \leq \|\mathbf{x}\|_2 \|A\mathbf{y}\|_2 \leq \|A\|_2 \|\mathbf{x}\|_2 \|\mathbf{y}\|_2. \qquad \square$$

1.2.1 Geometry

We start by looking at some basic definitions and results connected to geometry.

Definition 1.4. For a set $\mathcal{M} \subseteq \mathbb{R}^n$ we define:

- *Span* of $\mathcal{M}$, $\operatorname{span} \mathcal{M} := \left\{ \sum_{i=1}^p \theta_i \mathbf{x}_i \;\middle|\; \begin{array}{l} p \in \mathbb{Z}_{++},\ \boldsymbol{\theta} \in \mathbb{R}^p, \\ \mathbf{x}_1, \ldots, \mathbf{x}_p \in \mathcal{M} \end{array} \right\}$.
  If $\mathcal{M} = \operatorname{span} \mathcal{M}$, then we say that $\mathcal{M}$ is an *linear space.*

- *Affine hull* of $\mathcal{M}$, $\operatorname{aff}\mathcal{M} := \left\{ \sum_{i=1}^{p} \theta_i \mathbf{x}_i \;\middle|\; \begin{array}{l} p \in \mathbb{Z}_{++}, \\ \boldsymbol{\theta} \in \mathbb{R}^p,\ \mathbf{e}^\mathsf{T}\boldsymbol{\theta} = 1, \\ \mathbf{x}_1, \ldots, \mathbf{x}_p \in \mathcal{M} \end{array} \right\}$.
  If $\mathcal{M} = \operatorname{aff}\mathcal{M}$, then we say that $\mathcal{M}$ is an *affine* set.

- *Convex hull* of $\mathcal{M}$, $\operatorname{conv}\mathcal{M} := \left\{ \sum_{i=1}^{p} \theta_i \mathbf{x}_i \;\middle|\; \begin{array}{l} p \in \mathbb{Z}_{++}, \\ \boldsymbol{\theta} \in \mathbb{R}^p_+,\ \mathbf{e}^\mathsf{T}\boldsymbol{\theta} = 1, \\ \mathbf{x}_1, \ldots, \mathbf{x}_p \in \mathcal{M} \end{array} \right\}$.
  If $\mathcal{M} = \operatorname{conv}\mathcal{M}$, then we say that $\mathcal{M}$ is a *convex* set.

- $\operatorname{cone}\mathcal{M} := \{\lambda\mathbf{x} \mid \lambda \in \mathbb{R}_+,\ \mathbf{x} \in \mathcal{M}\}$.
  If $\mathcal{M} = \operatorname{cone}\mathcal{M}$, then we say that $\mathcal{M}$ is a *cone* (or equivalently a *conic set*).

- *Conic hull* of $\mathcal{M}$, $\operatorname{conic}\mathcal{M} := \operatorname{cone}\operatorname{conv}\mathcal{M}$.

- *Closure* of $\mathcal{M}$, $\operatorname{cl}\mathcal{M} := \left\{ \mathbf{x} \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \text{for all } \varepsilon > 0\ \exists \mathbf{y}_\varepsilon \in \mathcal{M} \\ \text{such that } \|\mathbf{x} - \mathbf{y}_\varepsilon\|_2 \le \varepsilon \end{array} \right\}$.
  If $\mathcal{M} = \operatorname{cl}\mathcal{M}$, then we say that $\mathcal{M}$ is a *closed* set.

- *Interior* of $\mathcal{M}$, $\operatorname{int}\mathcal{M} := \left\{ \mathbf{x} \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \exists \varepsilon > 0 \text{ s.t. for all } \mathbf{y} \in \mathbb{R}^n \text{ with} \\ \|\mathbf{x} - \mathbf{y}\|_2 \le \varepsilon,\ \text{we have } \mathbf{y} \in \mathcal{M} \end{array} \right\}$.

- *Boundary* of $\mathcal{M}$, $\operatorname{bd}\mathcal{M} := \operatorname{cl}\mathcal{M} \setminus \operatorname{int}\mathcal{M}$.

- *Relative interior* of $\mathcal{M}$,
$$\operatorname{reint}\mathcal{M} := \left\{ \mathbf{x} \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \exists \varepsilon > 0 \text{ such that for all } \mathbf{y} \in \operatorname{aff}\mathcal{M} \\ \text{with } \|\mathbf{x} - \mathbf{y}\|_2 \le \varepsilon,\ \text{we have } \mathbf{y} \in \mathcal{M} \end{array} \right\}.$$

- *Relative boundary* of $\mathcal{M}$, $\operatorname{rbd}\mathcal{M} := \operatorname{cl}\mathcal{M} \setminus \operatorname{reint}\mathcal{M}$.

- *Recession cone* of $\mathcal{M}$, $\operatorname{recc}\mathcal{M} := \{\mathbf{x} \in \mathbb{R}^n \mid (\mathcal{M} + \operatorname{cone}\{\mathbf{x}\}) \subseteq \mathcal{M}\}$.

- If $\mathcal{M}$ is a nonempty linear space, then the *dimension* of $\mathcal{M}$, denoted $\dim\mathcal{M}$, is the cardinality of a basis of $\mathcal{M}$.

- If $\mathcal{M}$ is a nonempty affine space with $\mathbf{x} \in \mathcal{M}$, then the *dimension* of $\mathcal{M}$, denoted $\dim\mathcal{M}$, is the dimension of the linear space $(\mathcal{M} - \{\mathbf{x}\})$.

- If $\mathcal{M}$ is a nonempty set, then the *dimension* of $\mathcal{M}$, denoted $\dim\mathcal{M}$, is the dimension of $\operatorname{aff}\mathcal{M}$.

- If $\dim\mathcal{M} = \dim\mathbb{R}^n$, then we say that $\mathcal{M}$ is a *full-dimensional* set.

- If $\mathcal{M} \cap (-\mathcal{M}) \subseteq \{\mathbf{0}\}$, then we say that $\mathcal{M}$ is a *pointed* set.
- If $\mathcal{M}$ is a closed, convex, pointed, full-dimensional cone, then we say that it is a *proper cone*.
- If there exists $R \in \mathbb{R}$ such that $\|\mathbf{x}\|_2 \leq R$ for all $\mathbf{x} \in \mathcal{M}$, then we say that $\mathcal{M}$ is a *bounded* set.

The following theorems follow immediately from the definitions:

Theorem 1.5. *For all operations ν given in Definition 1.4, excluding "*dim*", and for all $\mathcal{M} \subseteq \mathbb{R}^n$, we have $\nu(\nu(\mathcal{M})) = \nu(\mathcal{M})$.*

Theorem 1.6. *The interior of the intersection of finitely many sets is equal to the intersection of their interiors, i.e.* $\operatorname{int}(\bigcap_i \mathcal{M}_i) = \bigcap_i (\operatorname{int} \mathcal{M}_i)$.

Theorem 1.7. *Let $\mathcal{X}$ be one of the following properties for a set: linear, affine, convex, conic, bounded, closed or pointed. Then the intersection of (possibly infinitely many) sets with property $\mathcal{X}$, also has property $\mathcal{X}$.*

Theorem 1.8. *Let $\mathcal{X}$ be one of the following properties for a set: linear, affine, convex, conic, bounded or full-dimensional. Then the Minkowski sum of finitely many sets with property $\mathcal{X}$, also has property $\mathcal{X}$.*

We have the following example which demonstrates why certain properties are excluded from Theorems 1.7 and 1.8:

Example 1.9. We consider the following sets in $\mathbb{R}^3$:

$$\begin{aligned}
\mathcal{M}_1 &= \{\mathbf{x} \in \mathbb{R}^3 \mid x_1 \geq 0,\ x_3 \geq 0,\ x_2 + x_3 \geq 0,\ x_1(x_2 + x_3) \geq x_3^2\} \\
&= \operatorname{cl}\operatorname{cone}\{\mathbf{x} \in \mathbb{R}^3 \mid x_1 = +1,\ x_3 \geq 0,\ x_2 + x_3 \geq x_3^2\}, \\
\mathcal{M}_2 &= \{\mathbf{x} \in \mathbb{R}^3 \mid x_1 \leq 0,\ x_3 \geq 0,\ x_2 + x_3 \geq 0,\ x_1(x_2 + x_3) \leq -x_3^2\} \\
&= \operatorname{cl}\operatorname{cone}\{\mathbf{x} \in \mathbb{R}^3 \mid x_1 = -1,\ x_3 \geq 0,\ x_2 + x_3 \geq x_3^2\}.
\end{aligned}$$

These can be seen to be proper cones and it can also be shown that

$$\begin{aligned}
\mathcal{M}_1 + \mathcal{M}_2 &= \{\mathbf{x} \in \mathbb{R}^3 \mid x_3 > 0,\ x_2 + x_3 > 0\} \cup \{\mathbf{x} \in \mathbb{R}^3 \mid x_3 = 0,\ x_2 \geq 0\}, \\
\mathcal{M}_1 \cap \mathcal{M}_2 &= \{\mathbf{x} \in \mathbb{R}^3 \mid x_1 = x_3 = 0,\ x_2 \geq 0\}.
\end{aligned}$$

We now note that $\mathcal{M}_1 \cap \mathcal{M}_2$ is not full-dimensional, and $\mathcal{M}_1 + \mathcal{M}_2$ is neither pointed nor closed.

In the following theorem we shall look at a case when the Minkowski sum of two closed sets is in fact closed.

Theorem 1.10 ([Roc70, Corollary 9.1.2]). *Let $\mathcal{M}_1, \mathcal{M}_2 \subseteq \mathbb{R}^n$ be nonempty closed convex sets such that* $\operatorname{recc} \mathcal{M}_1 \cap \operatorname{recc}(-\mathcal{M}_2) = \{\mathbf{0}\}$. *Then $\mathcal{M}_1 + \mathcal{M}_2$ is closed.*

Corollary 1.11. *Let $\mathcal{M}_1, \mathcal{M}_2 \subseteq \mathbb{R}^n$ be closed convex sets, with at least one of the sets being bounded. Then $\mathcal{M}_1 + \mathcal{M}_2$ is closed.*

Corollary 1.12. *Consider $p \in \mathbb{Z}_{++}$ and closed convex cones $\mathcal{K}_1, \ldots, \mathcal{K}_p \subseteq \mathbb{R}^n$ such that $(\sum_{i=1}^p \mathcal{K}_i)$ is pointed. Then $(\sum_{i=1}^p \mathcal{K}_i)$ is a closed convex cone.*

We finish this subsection by considering the following four well-known lemmas.

Lemma 1.13. *In the definitions of the affine hull and the convex hull, we may limit p to be less than or equal to* $\dim(\mathcal{M}) + 1$ *without altering the sets. Similarly, in the definition of the linear span, we may limit p to be less than or equal to* $\dim(\mathcal{M})$ *without altering the set.*

Proof. This follows directly from Carathéodory's theorem, see for example [Roc70, Section 17]. $\square$

Lemma 1.14. *Consider a closed bounded set $\mathcal{M} \subseteq \mathbb{R}^n$ and a continuous function f from $\mathcal{M}$ to $\mathbb{R}$. Then there exists $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{M}$ such that for all $\mathbf{x} \in \mathcal{M}$ we have $f(\mathbf{x}_1) \leq f(\mathbf{x}) \leq f(\mathbf{x}_2)$.*

Proof. This is a well-known result, see for example [Roy63, page 36]. $\square$

Lemma 1.15. *Consider a closed bounded set $\mathcal{M} \subseteq \mathbb{R}^m$ and a continuous function f from $\mathcal{M}$ to $\mathbb{R}^n$. Then the set $\mathcal{L} = \{f(\mathbf{x}) \mid \mathbf{x} \in \mathcal{M}\}$ is a closed bounded set.*

Proof. This follows trivially from Lemma 1.14 and the definition of a continuous function. $\square$

Lemma 1.16. *Consider a closed set $\mathcal{M} \subseteq \mathbb{R}^m$ and a continuous function f from $\mathbb{R}^n$ to $\mathbb{R}^m$. Then the set $\mathcal{L} = \{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) \in \mathcal{M}\}$ is closed.*

Proof. This follows trivially from the definitions of a closed set and a continuous function. $\square$

1.2.2 Mathematical Optimisation

In *Mathematical Optimisation*, we are typically given functions $f_0, \ldots, f_m$ from $\mathbb{R}^n$ to $\mathbb{R}$, and wish to find the infimum value of $f_0(\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^n$ such that $f_i(\mathbf{x}) \geq 0$ for all $i = 1, \ldots, m$. This is denoted as follows:

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \quad & f_0(\mathbf{x}) \\ \text{s.t.} \quad & f_i(\mathbf{x}) \geq 0. \qquad \text{for all } i = 1, \ldots, m \end{aligned} \tag{1.7}$$

The infimum value for this problem is referred to as the *optimal value* and denoted by Val (1.7). We further define the *feasible set*,

$$\mathrm{Feas}\,(1.7) = \{\mathbf{x} \in \mathbb{R}^n \mid f_i(\mathbf{x}) \geq 0 \text{ for all } i = 1, \ldots, m\},$$

and the set of *optimal solutions*,

$$\mathrm{Opt}\,(1.7) = \{\mathbf{x} \in \mathrm{Feas}\,(1.7) \mid f_0(\mathbf{x}) = \mathrm{Val}\,(1.7)\}.$$

Remark 1.17. It is standard in the field of mathematical optimisation to write "min" rather than "inf" and refer to the minimum rather than the infimum, in spite of the fact that we do not assume that Opt (1.7) $\neq \emptyset$. Similarly for "max" and "sup". We shall stick to this standard from now on.

If Opt (1.7) $\neq \emptyset$ then we say that the minimum is *attained*, and in the following lemma we look at a well-known sufficient condition for the minimum being attained.

Lemma 1.18. *Let f_0 be a continuous function and* Feas (1.7) *be a closed bounded non-empty set. Then* Val (1.7) *is finite and* Opt (1.7) $\neq \emptyset$*.*

Proof. This comes directly from Lemma 1.14. □

A special type of mathematical optimisation problem that we shall look at in this thesis is a *conic optimisation* problem. In this we are considering a problem in one of the following forms, where $\mathcal{K} \subseteq \mathbb{R}^n$ is a closed convex cone, $\mathbf{c}, \mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$ and $b_1, \ldots, b_m \in \mathbb{R}$:

$$\begin{aligned}
\min_{\mathbf{x} \in \mathbb{R}^n} \quad & \langle \mathbf{c}, \mathbf{x} \rangle & \\
\text{s.t.} \quad & \langle \mathbf{a}_i, \mathbf{x} \rangle = b_i & \text{for all } i = 1, \ldots, m \\
& \mathbf{x} \in \mathcal{K}, &
\end{aligned}$$

$$\begin{aligned}
\max_{\mathbf{y} \in \mathbb{R}^m} \quad & \sum_{i=1}^m y_i b_i \\
\text{s.t.} \quad & \mathbf{c} - \sum_{i=1}^m y_i \mathbf{a}_i \in \mathcal{K}.
\end{aligned}$$

Remark 1.19. The two forms are considered together as it is a trivial task to take a problem written in one of these forms and rewrite it in the alternative form (although the values of the parameters would change). For example, suppose that we were given a problem in the latter form. Without loss of generality, assume that the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$ are orthonormal. There exists vectors $\mathbf{a}_{m+1}, \dots, \mathbf{a}_n \in \mathbb{R}^n$ such that $\mathbf{a}_1, \dots, \mathbf{a}_n$ is an orthonormal basis of $\mathbb{R}^n$. Now let $\widehat{\mathbf{c}} \in \mathbb{R}^n$ and $\gamma, b_{m+1}, \dots, b_n \in \mathbb{R}$ be such that

$$\widehat{\mathbf{c}} = \sum_{i=1}^{m} b_i \mathbf{a}_i, \qquad \gamma = \langle \mathbf{c}, \widehat{\mathbf{c}} \rangle,$$
$$b_i = \langle \mathbf{c}, \mathbf{a}_i \rangle \qquad \text{for all } i = m+1, \dots, n.$$

Now, for an arbitrary $\mathbf{x} \in \mathbb{R}^n$, we have that $\langle \mathbf{a}_i, \mathbf{x} \rangle = b_i$ for all $i = m+1, \dots, n$ if and only if there exists $\mathbf{y} \in \mathbb{R}^m$ such that $\mathbf{x} = \mathbf{c} - \sum_{i=1}^m y_i \mathbf{a}_i$. For such a $\mathbf{y}$ we have $y_i = \langle \mathbf{c}, \mathbf{a}_i \rangle - \langle \mathbf{a}_i, \mathbf{x} \rangle$ for all $i = 1, \dots, m$, and thus $\sum_{i=1}^m y_i b_i = \gamma - \langle \widehat{\mathbf{c}}, \mathbf{x} \rangle$. Therefore the problem that we are considering is equivalent to the following:

$$\begin{aligned}
\max_{\mathbf{x} \in \mathbb{R}^n} \quad & \gamma - \langle \widehat{\mathbf{c}}, \mathbf{x} \rangle \\
\text{s.t.} \quad & \langle \mathbf{a}_i, \mathbf{x} \rangle = b_i \qquad \text{for all } i = m+1, \dots, n \\
& \mathbf{x} \in \mathcal{K}.
\end{aligned}$$

Remark 1.20. Historically the word "programming" has been used in the place of the word "optimisation". However, the recent trend is very much moving away from this. In fact, in 2010, one of the main optimisation societies changed its name from the "Mathematical Programming Society" to the "Mathematical Optimization Society". In keeping with this trend, we shall also refer to "optimisation" rather than "programming" in this thesis.

1.2.3 Duality

A highly important concept in the field of conic optimisation is that of duality, which we shall consider in this subsection.

We begin by defining the dual of a set.

Definition 1.21. For a set $\mathcal{M} \subseteq \mathbb{R}^n$, we define its dual as

$$\mathcal{M}^* = \{\mathbf{y} \in \mathbb{R}^n \mid \langle \mathbf{y}, \mathbf{x} \rangle \geq 0 \text{ for all } \mathbf{x} \in \mathcal{M}\}.$$

Connected to this set we have the following lemma.

Lemma 1.22 ([Ber73, Section 2]). *For $\mathcal{M} \subseteq \mathbb{R}^n$ we have that*

- *$\mathcal{M}^*$ is a closed convex cone,*
- *$\mathcal{M}^{**} = \mathcal{M}$ if and only if $\mathcal{M}$ is a closed convex cone.*

The importance of this set comes from considering the following pair of conic optimisation problems, where $\mathcal{K} \subseteq \mathbb{R}^n$ is a closed convex cone and

$\mathbf{c}, \mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$ and $b_1, \ldots, b_m \in \mathbb{R}$:

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \quad & \langle \mathbf{c}, \mathbf{x} \rangle \\ \text{s.t.} \quad & \langle \mathbf{a}_i, \mathbf{x} \rangle = b_i \qquad \text{for all } i = 1, \ldots, m \\ & \mathbf{x} \in \mathcal{K}, \end{aligned} \tag{P1.8}$$

$$\begin{aligned} \max_{\mathbf{y}\in\mathbb{R}^m} \quad & \sum_{i=1}^m y_i b_i \\ \text{s.t.} \quad & \mathbf{c} - \sum_{i=1}^m y_i \mathbf{a}_i \in \mathcal{K}^*. \end{aligned} \tag{D1.8}$$

We now consider the following well-known lemma connected to these problems.

Lemma 1.23. *For all* $\mathbf{x} \in \operatorname{Feas}(\text{P1.8})$ *and* $\mathbf{y} \in \operatorname{Feas}(\text{D1.8})$, *we have*

$$\langle \mathbf{c}, \mathbf{x} \rangle \geq \sum_{i=1}^m y_i b_i,$$

and thus $\operatorname{Val}(\text{P1.8}) \geq \operatorname{Val}(\text{D1.8})$.

Proof. For all $\mathbf{x} \in \operatorname{Feas}(\text{P1.8})$ and $\mathbf{y} \in \operatorname{Feas}(\text{D1.8})$, we have

$$\langle \mathbf{c}, \mathbf{x} \rangle - \sum_{i=1}^m y_i b_i = \langle \mathbf{c}, \mathbf{x} \rangle - \sum_{i=1}^m y_i \langle \mathbf{a}_i, \mathbf{x} \rangle = \left\langle \mathbf{c} - \sum_{i=1}^m y_i \mathbf{a}_i, \mathbf{x} \right\rangle \geq 0. \qquad \square$$

This property is referred to as *weak duality*, and due to this property, such a pair of problems are referred to as primal and dual problems. Of especial interest is when the optimal values of these problems are in fact equal, and when this happens we say that there is *strong duality*. The most commonly considered condition for this is *Slater's condition*, which is summed up in the following definition and two theorems. For more details, it is recommended to read [SS00, section 4.1.2].

Definition 1.24. For problems (P1.8) and (D1.8):

- If there exists an $\mathbf{x} \in \operatorname{Feas}(\text{P1.8}) \cap \operatorname{int} \mathcal{K}$, then we say that Slater's condition holds for problem (P1.8) and $\mathbf{x}$ is referred to as a *strictly feasible* point of (P1.8),
- If there exists a $\mathbf{y} \in \operatorname{Feas}(\text{D1.8})$ such that $(\mathbf{c} - \sum_{i=1}^m y_i \mathbf{a}_i) \in \operatorname{int} \mathcal{K}^*$, then we say that Slater's condition holds for problem (D1.8) and $\mathbf{y}$ is referred to as a *strictly feasible* point of (D1.8).

Theorem 1.25. *If Slater's condition holds for problem* (P1.8), *then we have* $\mathrm{Val}\,(\mathrm{P1.8}) = \mathrm{Val}\,(\mathrm{D1.8})$. *Further more, if* $\mathrm{Val}\,(\mathrm{P1.8})$ *is finite, then the set* $\mathrm{Opt}\,(\mathrm{D1.8})$ *is nonempty.*

Theorem 1.26. *If Slater's condition holds for problem* (D1.8), *then we have* $\mathrm{Val}\,(\mathrm{P1.8}) = \mathrm{Val}\,(\mathrm{D1.8})$. *Further more, if* $\mathrm{Val}\,(\mathrm{P1.8})$ *is finite, then the set* $\mathrm{Opt}\,(\mathrm{P1.8})$ *is nonempty.*

From now on in this subsection we shall consider further useful properties connected to duality.

Lemma 1.27 ([Ber73, Section 2])**.** *Consider sets* $\mathcal{M}_1, \mathcal{M}_2 \subseteq \mathbb{R}^n$. *Then we have that:*

- $\mathcal{M}_1 \subseteq \mathcal{M}_2$ *implies* $\mathcal{M}_1^* \supseteq \mathcal{M}_2^*$,
- $\mathcal{M}_1^{**} = \mathrm{cl}\,\mathrm{conic}\,\mathcal{M}_1$,
- $\mathcal{M}_1^* = (\mathrm{cl}\,\mathcal{M}_1)^* = (\mathrm{cone}\,\mathcal{M}_1)^* = (\mathrm{conv}\,\mathcal{M}_1)^*$.

Lemma 1.28 ([Ber73, Section 2] and [BV04, Section 2.6])**.** *Consider closed convex cones* $\mathcal{K}_1, \mathcal{K}_2 \subseteq \mathbb{R}^n$ *and* $\mathcal{K}_3 \subseteq \mathbb{R}^m$. *Then we have that:*

- $(\mathcal{K}_1 + \mathcal{K}_2)^* = \mathcal{K}_1^* \cap \mathcal{K}_2^*$,
- $(\mathcal{K}_1 \cap \mathcal{K}_2)^* = \mathrm{cl}(\mathcal{K}_1^* + \mathcal{K}_2^*)$,
- $(\mathcal{K}_1 \times \mathcal{K}_3)^* = \mathcal{K}_1^* \times \mathcal{K}_3^*$,
- $\mathcal{K}_1$ *is full dimensional if and only if* $\mathcal{K}_1^*$ *is pointed.*
- $\mathcal{K}_1$ *is pointed if and only if* $\mathcal{K}_1^*$ *is full dimensional.*
- $\mathcal{K}_1$ *is a proper cone if and only if* $\mathcal{K}_1^*$ *is a proper cone.*

We now look at the interior of the dual, along with some connected lemmas.

Lemma 1.29 ([Ber73, page 8])**.** *For a set* $\mathcal{M} \subseteq \mathbb{R}^n$ *we have that*

$$\mathrm{int}(\mathcal{M}^*) = \{\mathbf{y} \in \mathbb{R}^n \mid \langle \mathbf{x}, \mathbf{y} \rangle > 0 \textit{ for all } \mathbf{x} \in \mathcal{M} \setminus \{\mathbf{0}\}\}.$$

Lemma 1.30. *Let* $\mathcal{M} \subseteq \mathbb{R}^n$ *be a closed set with* $\mathrm{int}(\mathcal{M}^*) \neq \emptyset$. *Then for any* $\mathbf{y} \in \mathrm{int}(\mathcal{M}^*)$ *and* $\alpha \in \mathbb{R}$, *we have that the set* $\mathcal{L} = \{\mathbf{x} \in \mathcal{M} \mid \langle \mathbf{x}, \mathbf{y} \rangle \leq \alpha\}$ *is a closed bounded set.*

Proof. $\mathcal{L}$ is the intersection of closed sets, and thus from Theorem 1.7, we see that it is also closed.

If $\alpha < 0$, then from the definition of the dual, we have that $\mathcal{L} = \emptyset$ and we are done. From now on we shall consider when $\alpha \geq 0$.

Let $\mathcal{K} = \operatorname{cl}\operatorname{conic}\mathcal{M}$ and $\beta = \min\{\langle \mathbf{x}, \mathbf{y}\rangle \mid \mathbf{x} \in \mathcal{K},\ \|\mathbf{x}\|_2 = 1\}$. From Lemmas 1.18 and 1.29, we have that β is finite and positive. It can now be observed that

$$\begin{aligned}\max\{\|\mathbf{x}\|_2 \mid \mathbf{x} \in \mathcal{M},\ \langle \mathbf{x}, \mathbf{y}\rangle \leq \alpha\} &\leq \max\{\|\mathbf{x}\|_2 \mid \mathbf{x} \in \mathcal{K},\ \langle \mathbf{x}, \mathbf{y}\rangle \leq \alpha\} \\ &= \max\left\{\theta \;\middle|\; \begin{array}{l}\theta \in \mathbb{R}_+,\ \mathbf{z} \in \mathcal{K}, \\ \|\mathbf{z}\|_2 = 1,\ \theta\langle \mathbf{z}, \mathbf{y}\rangle \leq \alpha\end{array}\right\} \\ &= \frac{\alpha}{\beta}.\end{aligned}$$ □

Corollary 1.31. *Consider $\alpha \in \mathbb{R}_{++}$ and a closed convex cone $\mathcal{K} \subseteq \mathbb{R}^n$ such that $\operatorname{int}(\mathcal{K}^*) \neq \emptyset$. Then letting $\mathbf{y} \in \operatorname{int}(\mathcal{K}^*)$ and $\mathcal{B} = \{\mathbf{x} \in \mathcal{K} \mid \langle \mathbf{x}, \mathbf{y}\rangle = \alpha\}$, we have that*

- *$\mathcal{B}$ is a closed convex bounded set,*
- *$\mathcal{K} = \operatorname{cone}\mathcal{B}$,*
- *For all $\mathbf{x} \in \mathcal{K} \setminus \{0\}$, there exists a unique $\widehat{\mathbf{x}} \in \mathcal{B}$ and $\lambda \in \mathbb{R}_{++}$ such that $\mathbf{x} = \lambda\widehat{\mathbf{x}}$.*

In fact, the set $\mathcal{B}$ from the previous corollary is referred to as a *base* of $\mathcal{K}$.

We now finish this section by considering the duals of some commonly considered sets.

Theorem 1.32. *For $n \in \mathbb{Z}_{++}$, we have that*

$$\begin{array}{lll}(\mathbb{R}^n)^* = \{\mathbf{0}\}, & (\mathbb{R}^n_+)^* = \mathbb{R}^n_+, & (\mathcal{N}^n)^* = \mathcal{N}^n, \\ (\mathcal{S}^n_+)^* = \mathcal{S}^n_+, & (\mathcal{S}^n_+ \cap \mathcal{N}^n)^* = \mathcal{S}^n_+ + \mathcal{N}^n, & (\mathcal{C}^{*n})^* = \mathcal{C}^n.\end{array}$$

Proof. These results follow trivially from the definitions, with the exception of "$(\mathcal{S}^n_+ \cap \mathcal{N}^n)^* = \mathcal{S}^n_+ + \mathcal{N}^n$", for which Corollary 1.12 and Lemma 1.28 are also required. □

Theorem 1.33. *For $m, n \in \mathbb{Z}_{++}$, consider a closed convex cone $\mathcal{K} \subseteq \mathbb{R}^n$, and $A \in \mathbb{R}^{m\times n}$. Letting $\mathcal{L} = \{A\mathbf{x} \mid \mathbf{x} \in \mathcal{K}\}$ and $\mathcal{M} = \{\mathbf{u} \in \mathbb{R}^m \mid A^{\mathsf{T}}\mathbf{u} \in \mathcal{K}^*\}$, we have that:*

i. $\mathcal{L}$ and $\mathcal{M}$ are convex cones, with $\mathcal{M}$ also being closed,

ii. $\mathcal{L}^* = \mathcal{M}$ and $\mathcal{M}^* = \operatorname{cl}\mathcal{L}$, *iii. If there exists a* $\mathbf{y} \in \mathbb{R}^n$ *such that* $A^\mathsf{T}\mathbf{y} \in \operatorname{reint} \mathcal{K}^*$*, then* $\mathcal{L}$ *is closed.*

Proof. We shall prove each point in turn:

i. This follows trivially from the definitions and Lemma 1.16.

ii. This follows from noting that

$$\begin{aligned}\mathcal{L}^* &= \{\mathbf{u} \in \mathbb{R}^m \mid \langle \mathbf{u}, A\mathbf{x}\rangle \geq 0 \text{ for all } \mathbf{x} \in \mathcal{K}\}\\ &= \{\mathbf{u} \in \mathbb{R}^m \mid \langle A^\mathsf{T}\mathbf{u}, \mathbf{x}\rangle \geq 0 \text{ for all } \mathbf{x} \in \mathcal{K}\}\\ &= \{\mathbf{u} \in \mathbb{R}^m \mid A^\mathsf{T}\mathbf{u} \in \mathcal{K}^*\}\\ &= \mathcal{M},\end{aligned}$$

and $\mathcal{M}^* = \mathcal{L}^{**} = \operatorname{cl} \mathcal{L}$.

iii. If $\mathcal{K}$ is a linear space then $\mathcal{L}$ is also a linear space, and so it is closed. From now on we consider when this is not the case.

By transforming the coordinates, without loss of generality, we have $\mathcal{K} = \widehat{\mathcal{K}} \times \mathbb{R}^p$, where $\widehat{\mathcal{K}}$ is a closed convex pointed cone. We then correspondingly let $A = \begin{pmatrix} A_1 & A_2 \end{pmatrix}$ where $A_1 \in \mathbb{R}^{m\times(n-p)}$ and $A_2 \in \mathbb{R}^{m\times p}$. Letting $\mathbf{y}$ be as required in the lemma, we have

$$\begin{aligned}&\mathcal{K}^* = \widehat{\mathcal{K}}^* \times \{\mathbf{0}\}, && \operatorname{reint} \mathcal{K}^* = \operatorname{int} \widehat{\mathcal{K}}^* \times \{\mathbf{0}\},\\ &A_1^\mathsf{T}\mathbf{y} \in \operatorname{int} \mathcal{K}, && A_2^\mathsf{T}\mathbf{y} = \mathbf{0},\\ &\mathcal{L} = \{A_1\mathbf{u} + A_2\mathbf{v} \mid \mathbf{u} \in \widehat{\mathcal{K}},\ \mathbf{v} \in \mathbb{R}^p\}.\end{aligned}$$

For all $\alpha \in \mathbb{R}$ we have that

$$\begin{aligned}\{\mathbf{z} \in \mathcal{L} \mid \langle \mathbf{z}, \mathbf{y}\rangle \leq \alpha\} &= \left\{ A_1\mathbf{u} + A_2\mathbf{v} \;\middle|\; \begin{array}{l}\mathbf{u} \in \widehat{\mathcal{K}},\ \mathbf{v} \in \mathbb{R}^p,\\ \langle A_1\mathbf{u} + A_2\mathbf{v}, \mathbf{y}\rangle \leq \alpha\end{array}\right\}\\ &= \{A_1\mathbf{u} + A_2\mathbf{v} \mid \mathbf{u} \in \widehat{\mathcal{K}},\ \mathbf{v} \in \mathbb{R}^p,\ \langle \mathbf{u}, A_1^\mathsf{T}\mathbf{y}\rangle \leq \alpha\}\\ &= \left\{ A_1\mathbf{u} \;\middle|\; \begin{array}{l}\mathbf{u} \in \widehat{\mathcal{K}},\\ \langle \mathbf{u}, A_1^\mathsf{T}\mathbf{y}\rangle \leq \alpha\end{array}\right\} + \{A_2\mathbf{v} \mid \mathbf{v} \in \mathbb{R}^m\}.\end{aligned}$$

This is the sum of two closed convex sets and from Lemma 1.30, we see that the set $\{A_1\mathbf{u} \mid \mathbf{u} \in \widehat{\mathcal{K}},\ \langle \mathbf{u}, A_1^\mathsf{T}\mathbf{y}\rangle \leq \alpha\}$ is bounded. Therefore, from Corollary 1.11 we get that $\{\mathbf{z} \in \mathcal{L} \mid \langle \mathbf{z}, \mathbf{y}\rangle \leq \alpha\}$ is a closed set. As this is true for all $\alpha \in \mathbb{R}$, we get that $\mathcal{L}$ is closed, completing the proof. □

Corollary 1.34. *For* $m, n, p \in \mathbb{Z}_{++}$*, consider a closed convex cone* $\mathcal{K} \subseteq \mathbb{R}^p$*, and matrices* $A \in \mathbb{R}^{m\times n}$ *and* $B \in \mathbb{R}^{p\times n}$*, such that there exists a* $\widehat{\mathbf{x}} \in \mathbb{R}^n$ *with* $B\widehat{\mathbf{x}} \in \operatorname{reint} \mathcal{K}$*. Then letting* $\mathcal{L} = \{A\mathbf{x} \mid B\mathbf{x} \in \mathcal{K}\}$*, we have that*

$$\mathcal{L}^* = \{\mathbf{u} \mid \exists \mathbf{v} \in \mathcal{K}^* \text{ s.t. } A^\mathsf{T}\mathbf{u} = B^\mathsf{T}\mathbf{v}\}.$$

1.2.4 Special types of conic optimisation

In this subsection, we shall briefly review three special types of conic optimisation problems, namely linear optimisation, semidefinite optimisation and copositive optimisation. Another well-known type of conic optimisation problem is that of second order cone optimisation. However, this has not been included in the list below as we are not required to consider it in this thesis, and its inclusion would require the introduction of more definitions and results. Instead the paper [LVBL98] is recommend for a good introduction to this field.

Linear Optimisation

If in problems (P1.8) and (D1.8) we consider $\mathcal{K} = \mathbb{R}^n_+$, then these problems are referred to as a pair of primal-dual *linear optimisation* problems, where we note that it can be trivially seen that $\mathbb{R}^n_+$ is a proper cone and $(\mathbb{R}^n_+)^* = \mathbb{R}^n_+$.

A special fact about linear optimisation is that, provided at least one of the primal-dual problems is feasible, we get strong duality.

The most common algorithm used for solving this type of problem is the so called *simplex algorithm*. This works quite efficiently in practice, however all versions of the simplex algorithm so far suggested have exponential worst-case running time. There are also *interior point algorithms* which solve this type of problem in polynomial time [RTV97].

A good introduction to this subject is provided in [Dan98].

Semidefinite Optimisation

If in the equivalent forms of problems (P1.8) and (D1.8) for the space of symmetric matrices, we consider $\mathcal{K} = \mathcal{S}^n_+$, then these problems are referred to as a pair of primal-dual *semidefinite optimisation* problems, where we note that it can be trivially seen that $\mathcal{S}^n_+$ is a proper cone and $(\mathcal{S}^n_+)^* = \mathcal{S}^n_+$.

This type of problem can also be solved in polynomial time using interior point methods [NN93].

A good introduction to this subject is provided in [WSV00].

Copositive Optimisation

If in the equivalent forms of problems (P1.8) and (D1.8) for the space of symmetric matrices, we consider $\mathcal{K} = \mathcal{C}^n$, then these problems are referred to as a pair of primal-dual *copositive optimisation* problems, where we note that it can be trivially seen that the copositive cone is the dual of the completely positive cone and we shall see in Chapter 5 that both the copositive and completely positive cones are proper cones.

As we shall see in the following chapter, these types of problems are in general $\mathcal{NP}$-hard, and thus we would not expect to be able to solve them efficiently (unless $\mathcal{P} = \mathcal{NP}$). Instead we consider replacing the copositive and completely positive cones with approximations which we are able to efficiently optimise over. This shall be discussed further in Part III.

It is hoped that this thesis will provide a good introduction to this subject, however alternative introductions are provided by [Dür10, HUS10, Bom12, Bur12].

1.3 Maximum Weight Clique Problem

In this section, we will consider a copositive reformulation for the maximum weight clique problem. A *clique* in a simple graph is a subset of its vertices such that the graph induced by this subset is complete (i.e. all vertices in the subset are connected by an edge). In the *maximum weight clique problem*, weights are assigned to the vertices of the graph and we wish to find the clique of maximum total weight. This is a well-known $\mathcal{NP}$-hard problem, and is in fact an extension of the maximum clique problem, which was one of Karp's original 21 $\mathcal{NP}$-complete problems [Kar72].

Copositive positive reformulations of this problem have previously been studied in the literature. For example, in [MS65, Bom98] this problem was reformulated as a standard quadratic optimisation problem, then in [BDK$^+$00] standard quadratic optimisation problems were in turn reformulated as a copositive optimisation problems. In this chapter a new and more direct proof for this reformulation shall be presented.

Specifying the maximum weight clique problem more precisely, we consider an arbitrary graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, \dots, n\}$. We let the vertices have strictly positive weights such that w_i gives the weighting on the vertex i. For $\mathcal{J} \subseteq \mathcal{V}$ we then define $W(\mathcal{J}) := \sum_{i\in\mathcal{J}} w_i$. The value of the maximum weight clique then equals $\max\{W(\mathcal{J}) \mid \mathcal{J} \text{ is a clique in } G\}$.

In order to find a copositive reformulation of this, we need a few technical results.

Lemma 1.35. *Let* $\mathbf{d} \in \mathbb{R}^n_{++}$, $D = \mathrm{Diag}(\mathbf{d})$ *and* $\lambda \in \mathbb{R}$. *Then the following statements are equivalent:*

i. $\lambda \geq \sum_{i=1}^n d_i^{-1}$,

ii. $(\lambda D - E) \in \mathcal{S}_+$,

iii. $(\lambda D - E) \in \mathcal{C}$.

Proof. As $\mathcal{S}_+ \subseteq \mathcal{C}$, it is trivial to see that $(ii) \Rightarrow (iii)$.

We shall now show that $(iii) \Rightarrow (i)$.

To do this, we let $\mathbf{v} \in \mathbb{R}^n_{++}$ such that $v_i = d_i^{-1}$ for all i. We then have that $(\lambda D - E) \in \mathcal{C}^n$ implies $0 \leq \mathbf{v}^\mathsf{T}(\lambda D - E)\mathbf{v} = \left(\sum_{i=1}^n d_i^{-1}\right)\left(\lambda - \left(\sum_{i=1}^n d_i^{-1}\right)\right)$.

We now complete the proof by showing that $(i) \Rightarrow (ii)$. We begin by noting the following equality, which implies that $\left((\sum_{i=1}^n d_i^{-1})D - E\right) \in \mathcal{S}^n_+$,

$$\begin{aligned}
&\left(\textstyle\sum_{i=1}^n d_i^{-1}\right) D - E \\
&\quad = \sum_{i<j} \left(\sqrt{d_i d_j^{-1}}\, \boldsymbol{e}_i - \sqrt{d_j d_i^{-1}}\, \boldsymbol{e}_j\right)\left(\sqrt{d_i d_j^{-1}}\, \boldsymbol{e}_i - \sqrt{d_j d_i^{-1}}\, \boldsymbol{e}_j\right)^\mathsf{T}
\end{aligned}$$

Therefore, as $D \in \mathcal{S}^n_+$, we immediately see that for all $\lambda \geq \sum_{i=1}^n d_i^{-1}$, we have

$$\lambda D - E = \left(\left(\lambda - \textstyle\sum_{i=1}^n d_i^{-1}\right)D + \left((\textstyle\sum_{i=1}^n d_i^{-1})D - E\right)\right) \in \mathcal{S}^n_+. \qquad \square$$

Theorem 1.36. *Consider $A \in \mathcal{S}^n$ and let*

$$\mathfrak{J} = \left\{\mathcal{J} \subseteq \{1,\dots,n\} \;\middle|\; (A)_{ij} < \sqrt{|(A)_{ii}(A)_{jj}|} \text{ for all } i,j \in \mathcal{J} : i \neq j\right\} \setminus \{\emptyset\}.$$

For a nonempty index set $\mathcal{J} \subseteq \{1,\dots,n\}$, we let $A_\mathcal{J} \in \mathcal{S}^{|\mathcal{J}|}$ denote the principal submatrix of A corresponding to the index set $\mathcal{J}$. We then have

$$A \in \mathcal{C} \qquad \Leftrightarrow \qquad A_\mathcal{J} \in \mathcal{C} \text{ for all } \mathcal{J} \in \mathfrak{J}.$$

Proof. This can be proven by considering the copositive completion problem [HJR05], however we will briefly give an alternative direct proof.

The forward implication follows trivially from Theorem 1.1.

For the reverse implication we consider an arbitrary $A \notin \mathcal{C}$, and denote the support of a vector $\mathbf{z} \in \mathbb{R}^n$ by $\operatorname{support}(\mathbf{z}) = \{i \in \{1,\dots,n\} \mid (\mathbf{z})_i \neq 0\}$.

Let $\mathbf{x} \in \mathbb{R}^n_+$ such that $0 > \mathbf{x}^\mathsf{T}A\mathbf{x}$ with minimal support, i.e. there does not exist $\mathbf{y} \in \mathbb{R}^n_+$ with $0 > \mathbf{y}^\mathsf{T}A\mathbf{y}$ and $\operatorname{support}(\mathbf{y}) \subset \operatorname{support}(\mathbf{x})$. Letting $\mathcal{J} = \operatorname{support}(\mathbf{x})$, we have $A_\mathcal{J} \notin \mathcal{C}$, and we shall show that $\mathcal{J} \in \mathfrak{J}$, which will complete the proof.

Suppose for the sake of contradiction that $\mathcal{J} \notin \mathfrak{J}$. Then there exists $i,j \in \mathcal{J}$ such that $(A)_{ij} \geq \sqrt{|(A)_{ii}(A)_{jj}|}$ and $i \neq j$. Without loss of generality $|(A)_{ii}| \leq |(A)_{jj}|$ and we consider three cases:

i. $(A)_{ii} = 0$, and there exists $k \in \mathcal{J}$ such that $(A)_{ik} < 0$:

We have $k \neq i,j$, and letting $\mathbf{y} = \left((|(A)_{kk}| + 1)\boldsymbol{e}_i - 2(A)_{ik}\boldsymbol{e}_k\right) \in \mathbb{R}^n_+$, we get the contradiction

$$\operatorname{support}(\mathbf{y}) \subset \mathcal{J}, \qquad \mathbf{y}^\mathsf{T}A\mathbf{y} = -4(A)_{ik}^2\left(1 + |(A)_{kk}| - (A)_{kk}\right) < 0.$$

ii. $(A)_{ii} = 0$, and $(A)_{ik} \geq 0$ for all $k \in \mathcal{J}$.

This implies that $\mathbf{x}^\mathsf{T} A\boldsymbol{e}_i \geq 0$. Letting $\mathbf{y} = \big(\mathbf{x} - (\mathbf{x})_i \boldsymbol{e}_i\big) \in \mathbb{R}^n_+$, we get the contradiction

$$\operatorname{support}(\mathbf{y}) \subset \mathcal{J}, \qquad \mathbf{y}^\mathsf{T} A\mathbf{y} = \mathbf{x}^\mathsf{T} A\mathbf{x} - 2(\mathbf{x})_i \mathbf{x}^\mathsf{T} A\boldsymbol{e}_i < 0.$$

iii. $(A)_{ii} \neq 0$:

The proof of this case is an adaptation of that for [HP73, Lemma 3.1]. Without loss of generality $\sqrt{|(A)_{ii}|}(\mathbf{x})_i + \sqrt{|(A)_{jj}|}(\mathbf{x})_j = \sqrt{|(A)_{ii}(A)_{jj}|}$. For $\theta \in [0,1]$ we now define $\mathbf{y}_\theta \in \mathbb{R}^n_+$ such that for $k = 1, \ldots, n$ we have

$$(\mathbf{y}_\theta)_k = \begin{cases} \sqrt{|(A)_{jj}|}\;\theta & \text{if } k = i \\ \sqrt{|(A)_{ii}|}\;(1-\theta) & \text{if } k = j \\ (\mathbf{x})_k & \text{otherwise.} \end{cases}$$

We then have $\mathbf{x} \in \{\mathbf{y}_\theta \mid \theta \in [0,1]\}$.

Now considering $\mathbf{y}_\theta^\mathsf{T} A\mathbf{y}_\theta$, this is a quadratic polynomial in θ, and the coefficient on the θ^2 term equals $2\sqrt{|(A)_{ii}(A)_{jj}|}\big(\sqrt{|(A)_{ii}(A)_{jj}|} - (A)_{ij}\big)$, which is less than or equal to zero. Therefore $\mathbf{y}_\theta^\mathsf{T} A\mathbf{y}_\theta$ is a concave function, and when minimising it over a convex set, the minimum is attained at the boundary of the set. This now gives the contradiction

$$\begin{aligned} &\mathbf{y}_0, \mathbf{y}_1 \in \mathbb{R}^n_+, \qquad \operatorname{support}(\mathbf{y}_0) \subset \mathcal{J}, \qquad \operatorname{support}(\mathbf{y}_1) \subset \mathcal{J}, \\ &\min\{\mathbf{y}_0^\mathsf{T} A\mathbf{y}_0, \mathbf{y}_1^\mathsf{T} A\mathbf{y}_1\} = \min\{\mathbf{y}_\theta^\mathsf{T} A\mathbf{y}_\theta \mid \theta \in [0,1]\} \leq \mathbf{x}^\mathsf{T} A\mathbf{x} < 0. \end{aligned}$$ □

Now, returning to our original problem, we define the following set and consider a connected lemma.

$$\mathcal{B} = \left\{ B \in \mathcal{S}^n \;\middle|\; \begin{array}{ll} (B)_{ii} = 1/w_i & \text{for all } i \\ (B)_{ij} = 0 & \text{for all } (i,j) \in \mathcal{E} \\ (B)_{ij} \geq \sqrt{(B)_{ii}(B)_{jj}} & \text{for all } i \neq j,\ (i,j) \notin \mathcal{E} \end{array} \right\}.$$

Lemma 1.37. *For all $B \in \mathcal{B}$ and $\lambda \in \mathbb{R}$ such that $(\lambda B - E)_{kk} \geq 0$ for all k, we have*

$$(\lambda B - E)_{ij} < \sqrt{(\lambda B - E)_{ii}(\lambda B - E)_{jj}} \qquad \Leftrightarrow \qquad (i,j) \in \mathcal{E}.$$

Proof. If $(i,j) \in \mathcal{E}$, then $(\lambda B - E)_{ij} = -1$, and so the result holds.

If $(i,j) \notin \mathcal{E}$, then by noting that $\lambda \geq (B)_{kk}^{-1} = w_k > 0$ for all k, we get that

$$\begin{aligned} \sqrt{(\lambda B - E)_{ii}(\lambda B - E)_{jj}} &= \sqrt{\left(\lambda\sqrt{(B)_{ii}(B)_{jj}} - 1\right)^2 - \lambda\left(\sqrt{(B)_{ii}} - \sqrt{(B)_{jj}}\right)^2} \\ &\leq \lambda\sqrt{(B)_{ii}(B)_{jj}} - 1 \leq \lambda(B)_{ij} - 1 = (\lambda B - E)_{ij}. \end{aligned}$$ □

Combining this with Theorem 1.36, we get the following corollary.

Corollary 1.38. *For all $B \in \mathcal{B}$ and $\lambda \in \mathbb{R}$, we have that the following statements are equivalent:*

i. *$\lambda B - E$ is copositive,*

ii. *For all cliques $\mathcal{J} \subseteq \mathcal{V}$, we have that the principal submatrix of $\lambda B - E$ corresponding corresponding to this clique, denoted $(\lambda B - E)_{\mathcal{J}}$, is copositive.*

Remark 1.39. If $B \in \mathcal{B}$ and $\mathcal{J} \subseteq \mathcal{V}$ is a clique, then the principal submatrix matrix $(B)_{\mathcal{J}}$ is a diagonal matrix with strictly positive on-diagonal entries.

A copositive reformulation of the maximum weight clique problem now immediately follows.

Theorem 1.40. *For any $B \in \mathcal{B}$, we have that*

$$\begin{aligned}\max_{\mathcal{J}}\{W(\mathcal{J}) \mid \mathcal{J} \textit{ is a clique in } G\} &= \min_{\lambda\in\mathbb{R}}\{\lambda \mid \lambda B - E \in \mathcal{C}^n\} \\ &= \max_{X\in\mathcal{S}^n}\{\langle E, X\rangle \mid \langle B, X\rangle = 1,\ X \in \mathcal{C}^{*n}\}.\end{aligned}$$

Proof. We have that

$$\begin{aligned}\min_{\lambda\in\mathbb{R}}\{\lambda \mid \lambda B - E \in \mathcal{C}^n\} &= \min_{\lambda\in\mathbb{R}}\left\{\lambda \;\middle|\; \begin{array}{c}(\lambda B - E)_{\mathcal{J}} \in \mathcal{C}^{|\mathcal{J}|} \\ \text{for all cliques } \mathcal{J} \text{ in } G\end{array}\right\} \\ &\qquad \text{(Corollary 1.38)} \\ &= \min_{\lambda\in\mathbb{R}}\{\lambda \mid \lambda \geq W(\mathcal{J}) \text{ for all cliques } \mathcal{J} \text{ in } G\} \\ &\qquad \text{(Lemma 1.35 and Remark 1.39)} \\ &= \max_{\mathcal{J}}\{W(\mathcal{J}) \mid \mathcal{J} \text{ is a clique in } G\}.\end{aligned}$$

We now note that $B \in (\operatorname{int}(\mathcal{S}^n_+) + \mathcal{N}^n) \subseteq \operatorname{int}\mathcal{C}^n$. From this we see that the copositive reformulation has a strictly feasible point, and thus, from Slater's condition, we get equality with its dual, which is the completely positive reformulation. □

As the maximum weighted clique problem is an $\mathcal{NP}$-hard problem, we get that copositive optimisation is in general an $\mathcal{NP}$-hard problem.

A special case of the maximum weight clique problem is when all of the weights are equal to one. This is then referred to as the maximum clique problem. This problem is equivalent to finding the stability number of a graph, which shall be discussed below. This provides an important application of copositive optimisation which has previously been studied in [KP02].

A stable set in a simple graph (also referred to as an *independent set* or a *co-clique*) is a subset of its vertices such that there are no edges between any of the vertices in this subset. Note that this is simply a clique in the complement of the graph. The cardinality of the maximum stable set is then referred to as the *stability number* of the graph, which is denoted by $\alpha(G)$, and finding the value of this is an $\mathcal{NP}$-hard problem.

In order to consider this further, we let $A_G \in \mathcal{S}^n$ denote the adjacency matrix of a graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, \dots, n\}$. This is defined such that

$$(A_G)_{ij} = \begin{cases} 1 & \text{if } (i,j) \in \mathcal{E} \\ 0 & \text{otherwise.} \end{cases}$$

From the results in this section, we get the following result, which was originally shown in [KP02]:

$$\begin{aligned} \alpha(G) &= \min_{\lambda \in \mathbb{R}} \{\lambda \mid \lambda(I + A_G) - E \in \mathcal{C}^n\} \\ &= \max_{X \in \mathcal{S}^n} \{\langle E, X\rangle \mid \langle I + A_G, X\rangle = 1,\ X \in \mathcal{C}^{*n}\}. \end{aligned}$$

Chapter 2

Set-semidefinite Optimisation*

2.1 Set-semidefinite Cone

In this section we shall be looking at a generalisation of the copositive cone called the set-semidefinite cone. This was first introduced in the article [EJ08], and was then further studied in [EJ10, EP12, GS11].

For a set $\mathcal{M} \subseteq \mathbb{R}^n$, we define the cone

$$\mathcal{C}_{\mathcal{M}} := \{X \in \mathcal{S}^n \mid \mathbf{v}^{\mathsf{T}} X \mathbf{v} \geq 0 \text{ for all } \mathbf{v} \in \mathcal{M}\}$$

This set is called the $\mathcal{M}$*-semidefinite cone*, and, for a general $\mathcal{M}$, we refer to these sets as *set-semidefinite cones*. We thus have that the copositive cone can also be referred to as the $\mathbb{R}^n_+$-semidefinite cone and the positive semidefinite cone can be referred to as the $\mathbb{R}^n$-semidefinite cone.

As the $\mathcal{M}$-semidefinite cone is an intersection of closed convex cones and $\mathcal{S}^n_+ \subseteq \mathcal{C}_{\mathcal{M}}$, we get that $\mathcal{C}_{\mathcal{M}}$ is a closed convex full-dimensional cone. In [EP12], it was shown that the dual to this cone is given as follows, and we note that this is a closed convex pointed cone:

$$\begin{aligned}\mathcal{C}^*_{\mathcal{M}} &= \operatorname{cl}\operatorname{conic}\{\mathbf{v}\mathbf{v}^{\mathsf{T}} \mid \mathbf{v} \in \mathcal{M}\} \\ &= \operatorname{conv}\{\mathbf{v}\mathbf{v}^{\mathsf{T}} \mid \mathbf{v} \in \operatorname{cl}\operatorname{cone}\mathcal{M}\}.\end{aligned}$$

The set-semidefinite cones, and their duals, are useful as they extend on the applications of the copositive and completely positive cones, as we shall see in this chapter. Their similarity to the copositive and completely positive cones also has the advantage that theory from considering one of these pairs of cones can often be adapted for the other pair.

*Submitted as:

[DEP12] P.J.C. Dickinson, G. Eichfelder and J. Povh. Erratum to “On the set-semidefinite representation of nonconvex quadratic programs over arbitrary feasible sets” [Optim. Letters, 2012]. *Optimization Letters*, accepted.

2.2 Single Quadratic Constraint Problem

In this section we consider the following problem, where $Q, A \in \mathcal{S}^n$ and $\mathcal{K} \subseteq \mathbb{R}^n$ is a closed cone.

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \quad & \mathbf{x}^\mathsf{T} Q\mathbf{x} \\ \text{s.t.} \quad & \mathbf{x}^\mathsf{T} A\mathbf{x} = 1 \\ & \mathbf{x} \in \mathcal{K}. \end{aligned} \tag{P2.1}$$

We shall show that, under certain assumptions, the following problem is a reformulation of this.

$$\begin{aligned} \min_{X\in\mathcal{S}^n} \quad & \langle Q, X\rangle \\ \text{s.t.} \quad & \langle A, X\rangle = 1 \\ & X \in \mathcal{C}^*_{\mathcal{K}}. \end{aligned} \tag{Q2.1}$$

The assumptions that we shall consider are the following:

Assumption 2.1. *We have that* $\mathbf{x}^\mathsf{T} A\mathbf{x} \geq 0$ *for all* $\mathbf{x} \in \mathcal{K}$ *(i.e.* $A \in \mathcal{C}_{\mathcal{K}}$*).*

Assumption 2.2. *For all* $\mathbf{x} \in \mathcal{K}$ *such that* $\mathbf{x}^\mathsf{T} A\mathbf{x} = 0$*, we have* $\mathbf{x}^\mathsf{T} Q\mathbf{x} \geq 0$*.*

Assumption 2.3. *For all* $\Lambda \in \mathbb{R}_+$ *and* $\mathbf{x}, \mathbf{y} \in \mathcal{K}$ *such that* $\mathbf{x}^\mathsf{T} A\mathbf{x} = 1$ *and* $\mathbf{y}^\mathsf{T} A\mathbf{y} = 0$*, there exists* $\lambda \in \mathbb{R}$ *such that* $|\lambda| \geq \Lambda$ *and* $(\mathbf{x} + \lambda\mathbf{y}) \in \operatorname{Feas}(\text{P2.1})$*.*

Remark 2.4. If $A \in \mathcal{S}^n_+$, then Assumption 2.1 holds.

Remark 2.5. If $A \in \operatorname{int}(\mathcal{S}^n_+)$, then $\mathbf{y}^\mathsf{T} A\mathbf{y} = 0$ if and only if $\mathbf{y} = \mathbf{0}$, and thus Assumptions 2.2 and 2.3 hold.

Remark 2.6. If $A \in \mathcal{S}^n_+$, then $\mathbf{y}^\mathsf{T} A\mathbf{y} = 0$ if and only if $A\mathbf{y} = \mathbf{0}$. If we then also have that $\mathcal{K}$ is a convex cone, then Assumption 2.3 holds.

We now present the following results connected to these assumptions:

Theorem 2.7. *Consider problems* (P2.1) *and* (Q2.1) *such that either Assumption 2.1 holds or* $\operatorname{Feas}(\text{P2.1}) = \emptyset$ *(or both). Then we have*

$$\operatorname{Feas}(\text{Q2.1}) = \operatorname{conv}\left\{\mathbf{x}\mathbf{x}^\mathsf{T} \;\middle|\; \mathbf{x} \in \operatorname{Feas}(\text{P2.1})\right\} + \operatorname{conv}\left\{\mathbf{x}\mathbf{x}^\mathsf{T} \;\middle|\; \begin{array}{l} \mathbf{x} \in \mathcal{K}, \\ \mathbf{x}^\mathsf{T} A\mathbf{x} = 0 \end{array}\right\}.$$

Proof. It is trivial to see that

$$\operatorname{Feas}(\text{Q2.1}) \supseteq \operatorname{conv}\left\{\mathbf{x}\mathbf{x}^\mathsf{T} \;\middle|\; \mathbf{x} \in \operatorname{Feas}(\text{P2.1})\right\} + \operatorname{conv}\left\{\mathbf{x}\mathbf{x}^\mathsf{T} \;\middle|\; \begin{array}{l} \mathbf{x} \in \mathcal{K}, \\ \mathbf{x}^\mathsf{T} A\mathbf{x} = 0 \end{array}\right\}.$$

We shall now suppose that $\mathrm{Feas}\,(\mathrm{Q2.1}) \neq \emptyset$ and consider an arbitrary $X \in \mathrm{Feas}\,(\mathrm{Q2.1})$.

There exists a finite set $\{\mathbf{x}_1, \dots, \mathbf{x}_p\} \subseteq \mathcal{K}$ such that $X = \sum_{i=1}^{p} \mathbf{x}_i \mathbf{x}_i^\mathsf{T}$. We then have $1 = \langle A, X \rangle = \sum_{i=1}^{p} \mathbf{x}_i^\mathsf{T} A \mathbf{x}_i$.

If $\mathrm{Feas}\,(\mathrm{P2.1}) = \emptyset$, then as $\mathcal{K}$ is a cone, we have that $\mathbf{x}_i^\mathsf{T} A \mathbf{x}_i \leq 0$ for all i. This then gives a contradiction, and thus $\mathrm{Feas}\,(\mathrm{Q2.1}) = \emptyset$ and we are done.

From now on we consider when $\mathrm{Feas}\,(\mathrm{P2.1}) \neq \emptyset$ and Assumption 2.1 holds. We then have that $\mathbf{x}_i^\mathsf{T} A \mathbf{x}_i \geq 0$ for all i, and, without loss of generality, we may assume that there exists a $q \in \{1, \dots, p\}$ such that $\mathbf{x}_i^\mathsf{T} A \mathbf{x}_i > 0$ if and only if $i \in \{1, \dots, q\}$. We now let $\{\mathbf{y}_1, \dots, \mathbf{y}_p\} \subseteq \mathcal{K}$ and $\{\theta_1, \dots, \theta_p\} \subseteq \mathbb{R}_{++}$ be such that

$$
\begin{aligned}
\theta_i &= \mathbf{x}_i^\mathsf{T} A \mathbf{x}_i && \text{for all } i = 1, \dots, q, \\
\theta_i &= (p-q)^{-1} && \text{for all } i = q+1, \dots, p, \\
\mathbf{y}_i &= \frac{1}{\sqrt{\theta_i}} \mathbf{x}_i && \text{for all } i = 1, \dots, p.
\end{aligned}
$$

We then have

$$
\begin{aligned}
1 &= \langle A, X \rangle = \sum_{i=1}^{q} \mathbf{x}_i^\mathsf{T} A \mathbf{x}_i = \sum_{i=1}^{q} \theta_i, \\
1 &= \sum_{i=q+1}^{p} \theta_i, \\
\mathbf{y}_i^\mathsf{T} A \mathbf{y}_i &= \frac{1}{\theta_i} \mathbf{x}_i^\mathsf{T} A \mathbf{x}_i = \begin{cases} 1 & \text{for all } i = 1, \dots, q, \\ 0 & \text{for all } i = q+1, \dots, p, \end{cases}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\left(\sum_{i=1}^{q} \theta_i \mathbf{y}_i \mathbf{y}_i^\mathsf{T} \right) &\in \mathrm{conv} \left\{ \mathbf{x}\mathbf{x}^\mathsf{T} \;\middle|\; \mathbf{x} \in \mathrm{Feas}\,(\mathrm{P2.1}) \right\}, \\
\left(\sum_{i=q+1}^{p} \theta_i \mathbf{y}_i \mathbf{y}_i^\mathsf{T} \right) &\in \mathrm{conv} \left\{ \mathbf{x}\mathbf{x}^\mathsf{T} \;\middle|\; \begin{array}{l} \mathbf{x} \in \mathcal{K}, \\ \mathbf{x}^\mathsf{T} A \mathbf{x} = 0 \end{array} \right\},
\end{aligned}
$$

We now note that $X = \sum_{i=1}^{q} \theta_i \mathbf{y}_i \mathbf{y}_i^\mathsf{T} + \sum_{i=q+1}^{p} \theta_i \mathbf{y}_i \mathbf{y}_i^\mathsf{T}$, which completes the proof. $\square$

Corollary 2.8. *Consider problems* (P2.1) *and* (Q2.1) *with* $\mathrm{Feas}\,(\mathrm{P2.1}) = \emptyset$. *Then* $\mathrm{Val}\,(\mathrm{Q2.1}) = \mathrm{Val}\,(\mathrm{P2.1}) = \infty$ *and the optimal sets of both problems are empty.*

Theorem 2.9. *Consider problems* (P2.1) *and* (Q2.1) *such that* Feas (P2.1) $\neq \emptyset$, *Assumption 2.1 holds and at least one of Assumptions 2.2 or 2.3 holds. Then* Val (Q2.1) = Val (P2.1) *and*

$$\mathrm{Opt}\,(\mathrm{Q2.1}) = \mathrm{conv}\left\{\mathbf{x}\mathbf{x}^\mathsf{T} \;\middle|\; \mathbf{x} \in \mathrm{Opt}\,(\mathrm{P2.1})\right\} + \mathrm{conv}\left\{\mathbf{x}\mathbf{x}^\mathsf{T} \;\middle|\; \begin{array}{l} \mathbf{x} \in \mathcal{K}, \\ \mathbf{x}^\mathsf{T} A\mathbf{x} = 0, \\ \mathbf{x}^\mathsf{T} Q\mathbf{x} = 0 \end{array}\right\}.$$

Proof. If Assumption 2.2 holds then from Theorem 2.7, we immediately get the required result.

We now consider when Assumption 2.2 does not hold, but Assumption 2.3 does. There exists $\mathbf{y} \in \mathcal{K}$ such that $\mathbf{y}^\mathsf{T} A\mathbf{y} = 0$ and $\mathbf{y}^\mathsf{T} Q\mathbf{y} < 0$. If we now consider an arbitrary $\mathbf{x} \in$ Feas (P2.1), then from Assumption 2.3 we see that for all $\Lambda \in \mathbb{R}_+$ there exists $\lambda \in \mathbb{R}$ such that $|\lambda| \geq \Lambda$ and $(\mathbf{x}+\lambda\mathbf{y}) \in$ Feas (P2.1). Considering the limit as Λ tends to infinity implies that Val (P2.1) $= -\infty$ and Opt (P2.1) $= \emptyset$. From Theorem 2.7 we get that Val (Q2.1) $\leq$ Val (P2.1), which then implies that Val (Q2.1) $= -\infty$ and Opt (Q2.1) $= \emptyset$ as required. □

2.3 Standard Quadratic Optimisation Problem

The results from the previous section can be immediately applied to the completely positive cone. A case of special interest is the standard quadratic optimisation problem, which is defined to be a problem of the following form, where $Q \in \mathcal{S}^n$.

$$\begin{array}{ll} \min & \mathbf{x}^\mathsf{T} Q\mathbf{x} \\ \text{s.t.} & \mathbf{e}^\mathsf{T}\mathbf{x} = 1 \\ & \mathbf{x} \in \mathbb{R}^n_+ \end{array}$$

A good survey on this problem is provided by [Bom98], and it was shown in the ground breaking paper [BDK$^+$00] that the following optimisation problem is a reformulation of this:

$$\begin{array}{ll} \min & \langle Q, X\rangle \\ \text{s.t.} & \langle E, X\rangle = 1 \\ & X \in \mathcal{C}^{*n}. \end{array}$$

We now note that this also comes directly from the results in the previous section by letting $A = E$ and $\mathcal{K} = \mathbb{R}^n_+$.

2.4 Quadratic Binary Optimisation

In this section we consider the following two problems, where $\mathcal{M} \subseteq \mathbb{R}^n$ and $\mathcal{B} \subseteq \{1, \dots, n\}$ are closed sets, $Q \in \mathcal{S}^n$ is a symmetric matrix, and $\mathbf{b} \in \mathbb{R}^m$ and $\mathbf{c}, \mathbf{a}_1, \dots, \mathbf{a}_m \in \mathbb{R}^n$ are vectors:

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \quad & \mathbf{x}^\mathsf{T} Q\mathbf{x} + 2\mathbf{q}^\mathsf{T}\mathbf{x} \\ \text{s.t.} \quad & \mathbf{a}_i^\mathsf{T}\mathbf{x} = b_i && \text{for all } i = 1, \dots, m \\ & x_j \in \{0, 1\} && \text{for all } j \in \mathcal{B} \\ & \mathbf{x} \in \mathcal{M}, \end{aligned} \tag{P2.2}$$

$$\begin{aligned} \min_{(\mathbf{x},X)\in\mathbb{R}^n\times\mathcal{S}^n} \quad & \langle Q, X\rangle + 2\mathbf{q}^\mathsf{T}\mathbf{x} \\ \text{s.t.} \quad & \mathbf{a}_i^\mathsf{T}\mathbf{x} = b_i && \text{for all } i = 1, \dots, m \\ & \mathbf{a}_i^\mathsf{T} X\mathbf{a}_i = b_i^2 && \text{for all } i = 1, \dots, m \\ & x_j = X_{jj} && \text{for all } j \in \mathcal{B} \\ & \begin{pmatrix} 1 & \mathbf{x}^\mathsf{T} \\ \mathbf{x} & X \end{pmatrix} \in \mathcal{C}^*_{\{1\}\times\mathcal{M}}. \end{aligned} \tag{Q2.2}$$

Remark 2.10. Recall that $\mathcal{C}^*_{\{1\}\times\mathcal{M}} = \operatorname{conv}\{\mathbf{v}\mathbf{v}^\mathsf{T} \in \operatorname{cl}\operatorname{cone}(\{1\} \times \mathcal{M})\}$.

An important assumption connected to this problem is as follows:

Assumption 2.11. *For all $\mathbf{x} \in \mathcal{M}$ such that $\mathbf{a}_i^\mathsf{T}\mathbf{x} = b_i$ for all $i = 1, \dots, m$, we have that $(\mathbf{x})_j \in [0, 1]$ for all $j \in B$.*

(This assumption can be made to hold by either adding the constraints directly to $\mathcal{M}$ or by adding slack variables with nonnegative constraints.)

We begin our analysis of this problem by considering the asymptotic cone, where for a nonempty set $\mathcal{M} \subseteq \mathbb{R}^n$, the asymptopic cone is defined as follows:

$$\mathcal{M}_\infty := \left\{ \mathbf{x} \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \exists\{(\lambda_i, \mathbf{x}_i) \mid i \in \mathbb{Z}_+\} \subseteq \mathbb{R}_+ \times \mathcal{M} \\ \text{such that } \lim_{i\to\infty} \lambda_i = 0, \quad \lim_{i\to\infty} \lambda_i\mathbf{x}_i = \mathbf{x} \end{array} \right\}.$$

We note that this is always a closed cone and $\operatorname{recc}\mathcal{M} \subseteq \mathcal{M}_\infty \subseteq \operatorname{cl}\operatorname{cone}\mathcal{M}$. We also have the following lemma from [AT02], where they used an equivalent definition of the asymptotic cone.

Lemma 2.12. *[AT02, Lemma 2.1.1] Let $\mathcal{M} \subseteq \mathbb{R}^n$ be a nonempty closed set. Then*

$$\operatorname{cl}\operatorname{cone}(\{1\} \times \mathcal{M}) = (\{0\} \times \mathcal{M}_\infty) \cup \operatorname{cone}(\{1\} \times \mathcal{M}).$$

Therefore, if $\mathcal{M}$ is a closed cone, then $\operatorname{cl}\operatorname{cone}(\{1\} \times \mathcal{M}) = \mathbb{R}_+ \times \mathcal{M}$.

Burer in [Bur09] considered problem (P2.2) for $\mathcal{M} = \mathbb{R}^n_+$. He showed that in this case, provided Assumption 2.11 holds, we have that problem (Q2.2) is a reformulation of problem (P2.2). (Note that in this case $\mathcal{C}^*_{\{1\}\times\mathcal{M}} = \mathcal{C}^{*(n+1)}$.)

In [DEP12], this result was extended for more general $\mathcal{M}$ (where this paper in fact corrected results from the paper [EP12]). We will now consider the results in this paper using our new results from Section 2.2.

We begin by considering the following two technical lemmas.

Lemma 2.13. *Consider $A_1, \ldots, A_l \in \mathcal{S}^{n+1}$ and a closed cone $\mathcal{L} \subseteq \mathbb{R}^{n+1}$. Suppose that for all $q \in \{1, \ldots, l\}$ and $\mathbf{z} \in \mathcal{L}$, we have that*

$$\mathbf{z}^\mathsf{T} A_i \mathbf{z} = 0 \textit{ for all } i = 1, \ldots, q-1 \qquad \Rightarrow \qquad \mathbf{z}^\mathsf{T} A_q \mathbf{z} \geq 0.$$

Then for $\mathcal{K} = \{\mathbf{z} \in \mathcal{L} \mid \mathbf{z}^\mathsf{T} A_i \mathbf{z} = 0 \textit{ for all } i\}$, we have

$$\mathcal{C}^*_\mathcal{K} = \{Z \in \mathcal{C}^*_\mathcal{L} \mid \langle A_i, Z\rangle = 0 \textit{ for all } i\}.$$

Proof. It is trivial to see that $\mathcal{C}^*_\mathcal{K} \subseteq \{Z \in \mathcal{C}^*_\mathcal{L} \mid \langle A_i, Z\rangle = 0 \text{ for all } i\}$. We now consider an arbitrary $Z \in \mathcal{C}^*_\mathcal{L}$ such that $\langle A_i, Z\rangle = 0$ for all i. There exists a finite set $\{\mathbf{z}_1, \ldots, \mathbf{z}_M\} \subseteq \mathcal{L}$ such that $Z = \sum_{j=1}^M \mathbf{z}_j \mathbf{z}_j^\mathsf{T}$. We shall now show by induction on p that for all $p = 0, \ldots, l$ we have

$$\mathbf{z}_j^\mathsf{T} A_i \mathbf{z}_j = 0 \qquad \text{for all } i = 1, \ldots, p,\ j = 1, \ldots, M,$$

which would imply that $Z \in \mathcal{C}^*_\mathcal{K}$, and thus complete the proof.

The statement is trivially true for $p = 0$. Now, for the sake of induction, we assume that it is true for $p = q-1$, where $q \in \{1, \ldots, l\}$. From the conditions in the lemma, this implies that $\mathbf{z}_j^\mathsf{T} A_q \mathbf{z}_j \geq 0$ for all j. We now note that $0 = \langle A_q, Z\rangle = \sum_{j=1}^M \mathbf{z}_j^\mathsf{T} A_q \mathbf{z}_j$, which implies that $\mathbf{z}_j^\mathsf{T} A_q \mathbf{z}_j = 0$ for all j. Therefore the statement is also true for $p = q$, completing the proof. □

Lemma 2.14. *Consider $A_1, \ldots, A_l \in \mathcal{S}^{n+1}$, a closed cone $\mathcal{L} \subseteq \mathbb{R}^{n+1}$ and $\widehat{\mathbf{a}}_1, \ldots, \widehat{\mathbf{a}}_m \in \mathbb{R}^{n+1}$. Suppose that for all $q \in \{1, \ldots, l\}$ and $\mathbf{z} \in \mathcal{L}$ with $\widehat{\mathbf{a}}_j^\mathsf{T}\mathbf{z} = 0$ for all j we have that*

$$\mathbf{z}^\mathsf{T} A_i \mathbf{z} = 0 \textit{ for all } i = 1, \ldots, q-1 \qquad \Rightarrow \qquad \mathbf{z}^\mathsf{T} A_q \mathbf{z} \geq 0.$$

We now let

- *$\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$ and $b_1, \ldots, b_m \in \mathbb{R}$ be such that $\widehat{\mathbf{a}}_i^\mathsf{T} = \begin{pmatrix} -b_i & \mathbf{a}_i^\mathsf{T}\end{pmatrix}$ for all i.*
- *$k = \dim\{\mathbf{z} \in \mathbb{R}^{n+1} \mid \widehat{\mathbf{a}}_i^\mathsf{T}\mathbf{z} = 0 \textit{ for all } i\}$ and $B \in \mathbb{R}^{(n+1)\times k}$ such that*

$$\{B\mathbf{y} \mid \mathbf{y} \in \mathbb{R}^k\} = \{\mathbf{z} \in \mathbb{R}^{n+1} \mid \widehat{\mathbf{a}}_i^\mathsf{T}\mathbf{z} = 0 \textit{ for all } i\},$$

- $\mathcal{K} = \{\mathbf{z} \in \mathcal{L} \mid \widehat{\mathbf{a}}_i^\mathsf{T}\mathbf{z} = 0,\ \mathbf{z}^\mathsf{T} A_j \mathbf{z} = 0 \textit{ for all } i, j\}$.

Then we have

$$\mathcal{C}_{\mathcal{K}}^* = \left\{ Z \in \mathcal{C}_{\mathcal{L}}^* \;\middle|\; \begin{array}{l} \sum_i \widehat{\mathbf{a}}_i^\mathsf{T} Z \widehat{\mathbf{a}}_i = 0, \\ \langle A_j, Z\rangle = 0 \textit{ for all } j \end{array} \right\} \tag{2.3}$$

$$= \left\{ Z \in \mathcal{C}_{\mathcal{L}}^* \;\middle|\; \begin{array}{r} Z\widehat{\mathbf{a}}_i = \mathbf{0} \textit{ for all } i, \\ \langle A_j, Z\rangle = 0 \textit{ for all } j \end{array} \right\} \tag{2.4}$$

$$= \left\{ BYB^\mathsf{T} \in \mathcal{C}_{\mathcal{L}}^* \;\middle|\; Y \in \mathcal{S}^k,\ \langle B^\mathsf{T} A_j B, Y\rangle = 0 \textit{ for all } j \right\} \tag{2.5}$$

$$= \left\{ \widehat{X} = \begin{pmatrix} x & \mathbf{x}^\mathsf{T} \\ \mathbf{x} & X \end{pmatrix} \in \mathcal{C}_{\mathcal{L}}^* \;\middle|\; \begin{array}{r} \mathbf{a}_i^\mathsf{T}\mathbf{x} = b_i x \textit{ for all } i, \\ \mathbf{a}_i^\mathsf{T} X \mathbf{a}_i = b_i^2 x \textit{ for all } i, \\ \langle A_j, \widehat{X}\rangle = 0 \textit{ for all } j \end{array} \right\} \tag{2.6}$$

Proof. The characterisations (2.3),(2.4) and (2.5) come directly from combining Lemma 2.13 with the fact that $\mathcal{C}_{\mathcal{L}}^* \subseteq \mathcal{S}_+^{n+1}$ and basic properties of positive semidefinite matrices. Characterisation (2.6) can then seen to contain characterisation (2.4) and be contained in characterisation (2.3), which completes the proof. □

Using this lemma, it can now be seen that, provided Assumption 2.11 holds and we let

$$\widehat{Q} := \begin{pmatrix} 0 & \mathbf{q}^\mathsf{T} \\ \mathbf{q} & Q \end{pmatrix}, \qquad \widehat{\mathbf{a}}_i := \begin{pmatrix} -b_i \\ \mathbf{a}_i \end{pmatrix}, \qquad A_j := \begin{pmatrix} 0 & -\boldsymbol{e}_j^\mathsf{T} \\ -\boldsymbol{e}_j^\mathsf{T} & 2\boldsymbol{e}_j\boldsymbol{e}_j^\mathsf{T} \end{pmatrix} \quad \text{for all } i, j,$$

$$\mathcal{K} := \left\{ \mathbf{z} \in \operatorname{cl}\operatorname{cone}(\{1\} \times \mathcal{M}) \;\middle|\; \begin{array}{r} \widehat{\mathbf{a}}_i^\mathsf{T}\mathbf{z} = 0 \text{ for all } i = 1, \dots, m, \\ \mathbf{z}^\mathsf{T} A_j \mathbf{z} = 0 \text{ for all } j \in \mathcal{B} \end{array} \right\}$$
$$= \left\{ \begin{pmatrix} 0 \\ \mathbf{d} \end{pmatrix} \;\middle|\; \begin{array}{ll} \mathbf{d} \in \mathcal{M}_\infty, & \\ \mathbf{a}_i^\mathsf{T}\mathbf{d} = 0 & \forall i, \\ (\mathbf{d})_j = 0 & \forall j \in \mathcal{B} \end{array} \right\} \cup \operatorname{cone} \left\{ \begin{pmatrix} 1 \\ \mathbf{x} \end{pmatrix} \;\middle|\; \begin{array}{ll} \mathbf{x} \in \mathcal{M}, & \\ \mathbf{a}_i^\mathsf{T}\mathbf{x} = b_i & \forall i, \\ x_j \in \{0,1\} & \forall j \in \mathcal{B} \end{array} \right\},$$

(noting that $\mathcal{K}$ is a closed cone), then problems (P2.2) and (Q2.2) can be reformulated respectively as follows:

$$\begin{array}{ll} \min\limits_{\mathbf{z} \in \mathbb{R}^{n+1}} & \mathbf{z}^\mathsf{T} \widehat{Q} \mathbf{z} \\ \text{s.t.} & \mathbf{z}^\mathsf{T} E_{11} \mathbf{z} = 1 \\ & \mathbf{z} \in \mathcal{K}, \end{array} \tag{P2.7}$$

$$\begin{aligned} \min_{Z \in \mathcal{S}^{n+1}} \quad & \langle \widehat{Q}, Z \rangle \\ \text{s.t.} \quad & \langle E_{11}, Z \rangle = 1 \\ & Z \in \mathcal{C}_{\mathcal{K}}^*. \end{aligned} \tag{Q2.7}$$

Remark 2.15. We consider the formulation (2.6) of $\mathcal{C}_{\mathcal{K}}^*$ in (Q2.7) to give problem (Q2.2) as this is the standard form used. However, it may be more advantageous to use formulation (2.5), which would both reduce the number of variables and the number of constraints.

We can now use the results from Section 2.2 to consider when these problems are equal. To do this, we first introduce the following assumptions, where, using similar notation to that in [BJ10, Bur09, EP12, DEP12], we let

$$L_\infty := \{\mathbf{d} \in \mathcal{M}_\infty \mid \mathbf{a}_i^\mathsf{T}\mathbf{d} = 0 \text{ for all } i = 1, \ldots, m\}.$$

Assumption 2.16. *If there exists $\mathbf{x} \in \mathcal{M}$ such that $\mathbf{a}_i^\mathsf{T}\mathbf{x} = \mathbf{b}_i$ for all i, then $(\mathbf{d})_j = 0$ for all $\mathbf{d} \in L_\infty$, $j \in \mathcal{B}$.*

Assumption 2.17. *The optimal value of* (Q2.2) *is not equal to $-\infty$.*

Assumption 2.18. *For all $\mathbf{d} \in L_\infty$, $\mathbf{x} \in \mathrm{Feas}\,(\text{P2.2})$, $\Lambda \in \mathbb{R}_+$, there exists a $\lambda \in \mathbb{R}$ such that $|\lambda| \geq \Lambda$ and $\mathbf{x} + \lambda\mathbf{d} \in \mathrm{Feas}\,(\text{P2.2})$.*

We now present the following theorem, which was the main result from the paper [DEP12].

Theorem 2.19. *Consider problems* (P2.2) *and* (Q2.2) *and suppose that Assumption 2.11 holds and at least one of the following is true:*

i. $\mathrm{Feas}\,(\text{P2.2}) = \emptyset$

ii. *Assumptions 2.16 and 2.17 hold,*

iii. *Assumptions 2.16 and 2.18 hold,*

iv. *Assumption 2.17 holds and $\mathcal{B} = \emptyset$,*

v. *Assumption 2.18 holds and $\mathcal{B} = \emptyset$,*

vi. *$\mathcal{M}$ is bounded,*

vii. *For all $\mathbf{d} \in \mathcal{M}_\infty$, $\mathbf{x} \in \mathcal{M}$, $\Lambda \in \mathbb{R}_+$, there exists $\lambda \in \mathbb{R}$ such that $|\lambda| \geq \Lambda$ and $\mathbf{x} + \lambda\mathbf{d} \in \mathcal{M}$,*

viii. $\mathcal{M}_\infty = \mathrm{recc}(\mathcal{M})$,

ix. *$\mathcal{M}$ is convex.*

Then $\mathrm{Val}\,(\mathrm{Q2.2}) = \mathrm{Val}\,(\mathrm{P2.2})$ *and we have that*

$$\begin{aligned}&\left\{\begin{pmatrix}1 & \mathbf{x}^\mathsf{T}\\ \mathbf{x} & X\end{pmatrix} \;\middle|\; (\mathbf{x}, X) \in \mathrm{Feas}\,(\mathrm{Q2.2})\right\}\\ &= \mathrm{conv}\left\{\begin{pmatrix}1\\ \mathbf{x}\end{pmatrix}\begin{pmatrix}1\\ \mathbf{x}\end{pmatrix}^\mathsf{T} \;\middle|\; \mathbf{x} \in \mathrm{Feas}\,(\mathrm{P2.2})\right\} + \mathrm{conv}\left\{\begin{pmatrix}0\\ \mathbf{d}\end{pmatrix}\begin{pmatrix}0\\ \mathbf{d}\end{pmatrix}^\mathsf{T} \;\middle|\; \mathbf{d} \in L_\infty\right\},\end{aligned}$$

$$\begin{aligned}&\left\{\begin{pmatrix}1 & \mathbf{x}^\mathsf{T}\\ \mathbf{x} & X\end{pmatrix} \;\middle|\; (\mathbf{x}, X) \in \mathrm{Opt}\,(\mathrm{Q2.2})\right\}\\ &= \mathrm{conv}\left\{\begin{pmatrix}1\\ \mathbf{x}\end{pmatrix}\begin{pmatrix}1\\ \mathbf{x}\end{pmatrix}^\mathsf{T} \;\middle|\; \mathbf{x} \in \mathrm{Opt}\,(\mathrm{P2.2})\right\} + \mathrm{conv}\left\{\begin{pmatrix}0\\ \mathbf{d}\end{pmatrix}\begin{pmatrix}0\\ \mathbf{d}\end{pmatrix}^\mathsf{T} \;\middle|\; \begin{array}{l}\mathbf{d} \in L_\infty,\\ \mathbf{d}^\mathsf{T} Q\mathbf{d} = 0\end{array}\right\}.\end{aligned}$$

Proof. From reformulations (P2.7) and (Q2.7), and Theorems 2.7 and 2.9, we see that we get the required results provided Assumption 2.11 holds and either $\mathrm{Feas}\,(\mathrm{P2.2}) = \emptyset$, or Assumption 2.1 holds and at least one of Assumptions 2.2 or 2.3 hold.

We have thus covered the case when $\mathrm{Feas}\,(\mathrm{P2.2}) = \emptyset$, and from now on we consider when $\mathrm{Feas}\,(\mathrm{P2.2}) \neq \emptyset$.

As $E_{11} \in \mathcal{S}_+^{n+1}$, we have that Assumption 2.1 always holds.

As $\mathrm{Feas}\,(\mathrm{P2.2}) \neq \emptyset$, there exists an $\mathbf{x} \in \mathcal{M}$ such that $\mathbf{a}_i^\mathsf{T}\mathbf{x} = b_i$ for all $i = 1, \ldots, m$. Now using Lemma 2.14 we see that if Assumption 2.16 holds then

$$\begin{aligned}\left\{\mathbf{z} \in \mathcal{K} \;\middle|\; \mathbf{z}^\mathsf{T} E_{11}\mathbf{z} = 0\right\} &= \left\{\begin{pmatrix}0\\ \mathbf{d}\end{pmatrix} \in \mathbb{R}^{n+1} \;\middle|\; \begin{array}{l}\mathbf{d} \in \mathcal{M}_\infty,\\ \mathbf{a}_i^\mathsf{T}\mathbf{d} = 0 \text{ for all } i\\ (\mathbf{d})_j = 0 \text{ for all } j \in \mathcal{B}\end{array}\right\}\\ &= \left\{\begin{pmatrix}0\\ \mathbf{d}\end{pmatrix} \in \mathbb{R}^{n+1} \;\middle|\; \mathbf{d} \in L_\infty\right\}\end{aligned}$$

We will now go through each remaining point in turn:

ii. If Assumptions 2.16 and 2.17 hold then it can be seen that Assumption 2.2 holds.

iii. If Assumptions 2.16 and 2.18 hold then it can be seen that Assumption 2.3 holds.

iv,v. If $\mathcal{B} = \emptyset$ then Assumption 2.16 trivially holds.

vi. If $\mathcal{M}$ is bounded then $L_\infty \subseteq M_\infty = \{\mathbf{0}\}$, and thus Assumptions 2.16 and 2.18 also hold.

vii. If Assumption 2.11 holds and point (*vii*) is true, then it can be seen that Assumptions 2.16 and 2.18 also hold.

viii. If $\mathcal{M}_\infty = \operatorname{recc}(\mathcal{M})$ then, from the definitions, point (*vii*) is true.

ix. If $\mathcal{M}$ is convex then it is a well-known result that $\mathcal{M}_\infty = \operatorname{recc}(\mathcal{M})$. □

2.5 Considering a special case

In this section we will briefly consider the $\mathcal{M}$-semidefinite cone, for the special case of $\mathcal{M} = \mathcal{L} \times \mathbb{R}^q$, where $\emptyset \neq \mathcal{L} \subseteq \mathbb{R}^p$ and $p + q = n$. This is motivated in part by [NTZ11, Subsection 4.1], where they consider an application of this set for $\mathcal{L} = \mathbb{R}^p_+$.

Theorem 2.20. *For $\mathcal{M} = \mathcal{L} \times \mathbb{R}^q$, where $\emptyset \neq \mathcal{L} \subseteq \mathbb{R}^p$ and $p + q = n$, we have*

$$\mathcal{C}_\mathcal{M} = \operatorname{cl}\left\{ \begin{pmatrix} V & 0 \\ 0 & 0 \end{pmatrix} + W \;\middle|\; V \in \mathcal{C}_\mathcal{L},\ W \in \mathcal{S}^n_+ \right\},$$

$$\mathcal{C}^*_\mathcal{M} = \left\{ \begin{pmatrix} X & Z^\mathsf{T} \\ Z & Y \end{pmatrix} \in \mathcal{S}^n_+ \;\middle|\; X \in \mathcal{C}^*_\mathcal{L} \right\}.$$

Proof. From Lemma 1.28 we see that we need only prove the characterisation of $\mathcal{C}_\mathcal{M}$.

We have that $\mathcal{C}_\mathcal{M} = \mathcal{C}_{\operatorname{cone}(\mathcal{M})}$ and $\operatorname{cone}(\mathcal{M}) = \operatorname{cone}(\mathcal{L}) \times \mathbb{R}^q$. Thus from now on, without loss of generality, we shall assume that $\mathcal{L}$ is a cone.

We let

$$\mathcal{K} = \left\{ \begin{pmatrix} V & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} Q \\ I \end{pmatrix} Y \begin{pmatrix} Q \\ I \end{pmatrix}^\mathsf{T} \in \mathcal{S}^n \;\middle|\; V \in \mathcal{C}_\mathcal{L},\ Q \in \mathbb{R}^{p\times q},\ Y \in \mathcal{S}^q_+ \right\},$$

and note that

$$\mathcal{K} \subseteq \left\{ \begin{pmatrix} V & 0 \\ 0 & 0 \end{pmatrix} + W \;\middle|\; V \in \mathcal{C}_\mathcal{L},\ W \in \mathcal{S}^n_+ \right\} \subseteq \mathcal{C}_\mathcal{M}.$$

As was stated at the start of this chapter, $\mathcal{C}_\mathcal{M}$ is a closed convex full-dimensional cone. Therefore, if we can show that $\operatorname{int}\mathcal{C}_\mathcal{M} \subseteq \mathcal{K}$, then this would complete the proof.

We consider an arbitrary matrix $\begin{pmatrix} X & Z^\mathsf{T} \\ Z & Y \end{pmatrix} \in \operatorname{int}\mathcal{C}_\mathcal{M}$, where $X \in \mathcal{S}^p$, $Y \in \mathcal{S}^q$ and $Z \in \mathbb{R}^{q\times p}$. There exists a $\lambda > 0$ such that $\begin{pmatrix} X & Z^\mathsf{T} \\ Z & Y - \lambda I \end{pmatrix} \in \mathcal{C}_\mathcal{M}$. Therefore, due to $\mathcal{L}$ being a cone, and so $\mathbf{0} \in \mathcal{L}$, we get that for all $\mathbf{y} \in \mathbb{R}^q \setminus \{\mathbf{0}\}$

$$0 \leq \begin{pmatrix} \mathbf{0} \\ \mathbf{y} \end{pmatrix}^\mathsf{T} \begin{pmatrix} X & Z^\mathsf{T} \\ Z & Y - \lambda I \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \mathbf{y} \end{pmatrix} = \mathbf{y}^\mathsf{T} Y \mathbf{y} - \lambda \|\mathbf{y}\|_2^2 < \mathbf{y}^\mathsf{T} Y \mathbf{y}.$$

Therefore Y is positive definite, which implies that it has an inverse. We now let

$$\begin{aligned} Q &= Z^{\mathsf{T}} Y^{-1} \in \mathbb{R}^{p\times q}, \\ V &= (X - QYQ^{\mathsf{T}}) = (X - Z^{\mathsf{T}} Y^{-1} Z) \in \mathcal{S}^p. \end{aligned}$$

In fact, V is the Schur complement of the original matrix that we were considering. For all $\mathbf{u} \in \mathcal{L}$, we have that $-Q^{\mathsf{T}}\mathbf{u} \in \mathbb{R}^q$, and thus

$$0 \le \begin{pmatrix} \mathbf{u} \\ -Q^{\mathsf{T}}\mathbf{u} \end{pmatrix}^{\mathsf{T}} \begin{pmatrix} X & Z^{\mathsf{T}} \\ Z & Y \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ -Q^{\mathsf{T}}\mathbf{u} \end{pmatrix} = \mathbf{u}^{\mathsf{T}}(X - QYQ^{\mathsf{T}})\mathbf{u} = \mathbf{u}^{\mathsf{T}} V \mathbf{u}.$$

Therefore $V \in \mathcal{C}_{\mathcal{L}}$ and $Y \in \mathcal{S}^q_+$ and $Q \in \mathbb{R}^{p\times q}$ and

$$\begin{pmatrix} X & Z^{\mathsf{T}} \\ Z & Y \end{pmatrix} = \begin{pmatrix} V & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} Q \\ I \end{pmatrix} Y \begin{pmatrix} Q \\ I \end{pmatrix}^{\mathsf{T}} \in \mathcal{K}. \qquad \square$$

Remark 2.21. From Corollary 1.12 we get that if in the previous theorem $\mathcal{C}_{\mathcal{M}}$ was pointed, then the closure operation would have been unnecessary. Now, considering the paper [GS11], we get that a sufficient condition for this is that $\operatorname{int} \mathcal{M} \neq \emptyset$, or equivalently $\operatorname{int} \mathcal{L} \neq \emptyset$.

We can now reconsider Lemma 2.14 for this special case:

Corollary 2.22. *Consider a set $\mathcal{L} = \widehat{\mathcal{L}} \times \mathbb{R}^q$, where $p + q = n + 1$ and $\widehat{\mathcal{L}}$ is a nonempty closed cone in $\mathbb{R}^p$. We also consider vectors $\widehat{\mathbf{a}}_1, \ldots, \widehat{\mathbf{a}}_m \in \mathbb{R}^{n+1}$ and matrices $A_1, \ldots, A_l \in \mathcal{S}^{n+1}$ such that for all $q \in \{1, \ldots, l\}$ and $\mathbf{z} \in \mathcal{L}$ with $\mathbf{a}_i^{\mathsf{T}}\mathbf{z} = 0$ for all i we have that*

$$\mathbf{z}^{\mathsf{T}} A_i \mathbf{z} = 0 \text{ for all } i = 1, \ldots, q-1 \qquad \Rightarrow \qquad \mathbf{z}^{\mathsf{T}} A_q \mathbf{z} \ge 0.$$

We now let

- $k = \dim\{\mathbf{z} \in \mathbb{R}^{n+1} \mid \widehat{\mathbf{a}}_i^{\mathsf{T}}\mathbf{z} = 0 \text{ for all } i\}$,
- $B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}$, *where $B_1 \in \mathbb{R}^{p\times k}$ and $B_2 \in \mathbb{R}^{q\times k}$ such that*
$$\left\{ \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} \mathbf{y} \;\middle|\; \mathbf{y} \in \mathbb{R}^k \right\} = \{\mathbf{z} \in \mathbb{R}^{n+1} \mid \widehat{\mathbf{a}}_i^{\mathsf{T}}\mathbf{z} = 0 \text{ for all } i\},$$
- $\mathcal{K} = \{\mathbf{z} \in \mathcal{L} \mid \widehat{\mathbf{a}}_i^{\mathsf{T}}\mathbf{z} = 0,\ \mathbf{z}^{\mathsf{T}} A_i \mathbf{z} = 0 \text{ for all } i, j\}$.

Then we have

$$\mathcal{C}^*_{\mathcal{K}} = \left\{ BYB^{\mathsf{T}} \;\middle|\; \begin{array}{l} Y \in \mathcal{S}^k_+, \quad B_1 Y B_1^{\mathsf{T}} \in \mathcal{C}^*_{\widehat{\mathcal{L}}}, \\ \langle B^{\mathsf{T}} A_j B, Y\rangle = 0 \text{ for all } j \end{array} \right\}.$$

Note that in the previous corollary, that instead of a positive semidefinite constraint of order $(n+1)$, we fact only need a positive semidefinite constraint of order k.

Chapter 3

Complexity*

3.1 Membership Problems

In this chapter we shall look at the computational complexity of checking whether a matrix is copositive, or alternatively checking whether it is completely positive. In order to consider this, we start by looking at some definitions for general sets. We shall only give a brief overview here, and will not go in to detail about what we mean by encoding lengths and certificates. For more precise details, the book [GLS88] is recommended.

In this chapter we shall let $\mathcal{X}$ equal either $\mathbb{R}^n$ or $\mathcal{S}^n$, and correspondingly let $\mathcal{Q}$ equal either $\mathbb{Q}^n$ or $\mathbb{Q}^{n\times n}\cap\mathcal{S}^n$ respectively. Equivalent definitions to those given in this chapter can be found in [GLS88, Chapters 0 to 2].

Definition 3.1 (**Strong Membership Problem (MEM)).** Let $\mathcal{M} \subseteq \mathcal{X}$ and $Y \in \mathcal{Q}$. Then either

i. assert that $Y \in \mathcal{M}$, or

ii. assert that $Y \notin \mathcal{M}$.

Definition 3.2 (**Strong Separation Problem (SEP)).** Let $\mathcal{M} \subseteq \mathcal{X}$ and $Y \in \mathcal{Q}$. Then either

i. assert that $Y \in \mathcal{M}$, or

ii. give a $H \in \mathcal{Q}$ such that $\langle H, X\rangle > \langle H, Y\rangle$ for all $X \in \mathcal{M}$.

---

*Submitted as:

[DG11] P.J.C. Dickinson and L. Gijben. On the computational complexity of membership problems for the completely positive cone and its dual. *Computational Optimization and Applications*, conditionally accepted.

[ABD12] K.M. Anstreicher, S. Burer and P.J.C. Dickinson. An algorithm for computing the CP-factorization of a completely positive matrix. Under construction.

Definition 3.3 (Class $\mathcal{NP}$ for MEM). A strong membership problem is defined to be in the class $\mathcal{NP}$ if, whenever $Y \in \mathcal{M}$, there is a certificate of this which can be read and checked in polynomial time with respect to the encoding length of Y.

Definition 3.4 (Class co-$\mathcal{NP}$ for MEM). A strong membership problem is defined to be in the class co-$\mathcal{NP}$ if, whenever $Y \notin \mathcal{M}$, there is a certificate of this which can be read and checked in polynomial time with respect to the encoding length of Y.

We also consider the so called weak membership problem. To do this we first define the following sets:

Definition 3.5. Let $\mathcal{M} \subseteq \mathcal{X}$ and $\varepsilon > 0$. Then we define the ε outer and inner approximations of $\mathcal{M}$ respectively as follows:

$$\begin{aligned} \mathrm{S}(\mathcal{M}, \varepsilon) &:= \{X \in \mathcal{X} \mid \exists Y \in \mathcal{M} \text{ s.t. } \|X - Y\|_2 \leq \varepsilon\}, \\ \mathrm{S}(\mathcal{M}, -\varepsilon) &:= \{X \in \mathcal{X} \mid \mathrm{S}(\{X\}, \varepsilon) \subseteq \mathcal{M}\}. \end{aligned}$$

Furthermore, for $A \in \mathcal{S}^n$, we define $\mathrm{S}(A, \varepsilon) := \mathrm{S}(\{A\}, \varepsilon)$.

Note that for any monotonically decreasing sequence of positive scalars $\{\varepsilon_i \mid i \in \mathbb{Z}_+\}$, such that $\lim_{i\to\infty} \varepsilon_i = 0$, we have

$$\begin{aligned} \mathrm{S}(\mathcal{M}, -\varepsilon_0) \subseteq \mathrm{S}(\mathcal{M}, -\varepsilon_1) \subseteq \cdots \subseteq \textstyle\bigcup_{i\in\mathbb{Z}_+} \mathrm{S}(\mathcal{M}, -\varepsilon_i) &= \operatorname{int}\mathcal{M} \subseteq \mathcal{M}, \\ \mathrm{S}(\mathcal{M}, \varepsilon_0) \supseteq \mathrm{S}(\mathcal{M}, \varepsilon_1) \supseteq \cdots \supseteq \textstyle\bigcap_{i\in\mathbb{Z}_+} \mathrm{S}(\mathcal{M}, \varepsilon_i) &= \operatorname{cl}\mathcal{M} \supseteq \mathcal{M}, \end{aligned}$$

and thus $\mathrm{S}(\mathcal{M}, -\varepsilon)$ and $\mathrm{S}(\mathcal{M}, \varepsilon)$ can be seen as inner and outer approximations of $\mathcal{M}$ respectively.

We are now ready to define the weak membership problem.

Definition 3.6 (Weak Membership Problem (WMEM)). Let $\mathcal{M} \subseteq \mathcal{X}$, $Y \in \mathcal{Q}$ and $\varepsilon \in \mathbb{Q}_{++}$. Then assert either

i. $Y \in \mathrm{S}(\mathcal{M}, \varepsilon)$, or

ii. $Y \notin \mathrm{S}(\mathcal{M}, -\varepsilon)$.

Note that $\mathrm{S}(\mathcal{M}, \varepsilon) \setminus \mathrm{S}(\mathcal{M}, -\varepsilon) \supseteq \operatorname{cl}\mathcal{M} \setminus \operatorname{int}\mathcal{M} = \operatorname{bd}\mathcal{M}$, which is nonempty unless $\mathcal{M} \in \{\emptyset, \mathcal{X}\}$. Therefore, in general, for some values of Y and ε, either assertion would be valid.

The definitions of the classes $\mathcal{NP}$ and co-$\mathcal{NP}$ have also been extended to include this type of problem.

Definition 3.7 (The class $\mathcal{NP}$ for WMEM). A weak membership problem is defined to be in the class $\mathcal{NP}$ if, whenever $Y \in \mathrm{S}(\mathcal{M}, -\varepsilon)$, there is a certificate showing that $Y \in \mathrm{S}(\mathcal{M}, \varepsilon)$, which can be read and checked in polynomial time with respect to the encoding length of Y and ε. (That is to say, whenever Y is in the inner approximation, there is a certificate showing it to be in the outer approximation.)

Definition 3.8 (The class co-$\mathcal{NP}$ for WMEM). A weak membership problem is defined to be in the class co-$\mathcal{NP}$ if, whenever $Y \notin \mathrm{S}(\mathcal{M}, \varepsilon)$, there is a certificate showing that $Y \notin \mathrm{S}(\mathcal{M}, -\varepsilon)$, which can be read and checked in polynomial time with respect to the encoding length of Y and ε. (That is to say, whenever Y is not in the outer approximation, there is a certificate showing it not to be in the inner approximation.)

The following lemma now follows immediately from the definitions:

Lemma 3.9. *If a strong membership problem is in $\mathcal{NP}$ (co-$\mathcal{NP}$) then the corresponding weak membership problem is also in $\mathcal{NP}$ (co-$\mathcal{NP}$).*

Related to this we also have the following lemma. In this, we recall that a problem is defined to be $\mathcal{NP}$-hard if being able to solve it in polynomial time with respect to the encoding lengths of its inputs would allow us to solve all problems in the classical class $\mathcal{NP}$ (i.e. excluding the extension of $\mathcal{NP}$ for weak membership) in polynomial time with respect to the encoding lengths of their inputs. For a more thorough discussion of $\mathcal{NP}$-hard problems, the book [GJ79] is recommended.

Lemma 3.10. *If a weak membership problem is $\mathcal{NP}$-hard then the corresponding strong membership problem is also $\mathcal{NP}$-hard.*

One final related definition that we shall give in this section is that of quintuples:

Definition 3.11. Consider a closed convex set $\mathcal{M} \subseteq \mathcal{X}$ such that there exists an $A_0 \in \mathcal{Q}$ and $r, R \in \mathbb{Q}_{++}$ such that

$$\mathrm{S}(A_0, r) \subseteq \mathcal{M} \subseteq \mathrm{S}(0, R).$$

Then for $N = \dim \mathcal{X}$, we have that $(\mathcal{M}; N, R, r, A_0)$ is defined to be a valid quintuple for $\mathcal{M}$.

3.2 Ellipsoid Method

In this section we will give a quick sketch of the ellipsoid method. For more details [GLS88, Chapter 3] is recommended.

We consider the following optimisation problem, where $\mathcal{M} \subseteq \mathbb{R}^n$ is a closed convex set, with a valid quintuple $(\mathcal{M}; N, R, r, \mathbf{a}_0)$, and $\mathbf{q} \in \mathbb{Q}^n$:

$$\nu = \min_{\mathbf{x}\in\mathbb{R}^n} \{\langle \mathbf{q}, \mathbf{x}\rangle \mid \mathbf{x} \in \mathcal{M}\}.$$

For $\delta \in \mathbb{Q}_{++}$, we wish to find a $\widehat{\mathbf{x}} \in \mathcal{M}$ such that $\langle \mathbf{q}, \widehat{\mathbf{x}}\rangle \leq \nu + \delta$. We shall assume that $\delta \leq 2R\|\mathbf{q}\|_2$, otherwise for all $\mathbf{x} \in \mathcal{M}$, including $\mathbf{a}_0$, we would have that $\langle \mathbf{q}, \mathbf{x}\rangle \leq \nu + \delta$, and thus the problem would be trivial. We now note an important result connected to this problem.

Lemma 3.12. *Consider $\mathbf{q} \in \mathbb{R}^n$, a closed convex set $\mathcal{M} \subseteq \mathbb{R}^n$, with an associated valid quintuple $(\mathcal{M}; N, R, r, \mathbf{a}_0)$, and $\delta \in (0, 2R\|\mathbf{q}\|_2]$. Then we have that*

$$\mathrm{Vol}\left(\{\mathbf{x} \in \mathcal{M} \mid \langle \mathbf{q}, \mathbf{x}\rangle \leq \nu + \delta\}\right) \geq \left(r\delta/(2R\|\mathbf{q}\|_2)\right)^N \mathrm{Vol}\left(S(\mathbf{0}, 1)\right).$$

Proof. We shall prove the required result by showing that there is a ball of radius $t = r\delta/(2R\|\mathbf{q}\|_2)$ contained in the set $\{\mathbf{x} \in \mathcal{M} \mid \langle \mathbf{q}, \mathbf{x}\rangle \leq \nu + \delta\}$.

As $\mathcal{M}$ is a closed bounded set, from Lemma 1.18, we see that there exists $\mathbf{b} \in \mathcal{M}$ such that $\nu = \langle \mathbf{q}, \mathbf{b}\rangle$. We then have that

$$\mathrm{S}\left(\mathbf{b}, \delta/\|\mathbf{q}\|_2\right) \subseteq \{\mathbf{x} \in \mathbb{R}^n \mid |\langle \mathbf{q}, \mathbf{x}\rangle - \nu| \leq \delta\}.$$

As $\mathcal{M}$ is convex, for all $\theta \in [0, 1]$ we have $\mathrm{S}\left(\theta\mathbf{a}_0 + (1-\theta)\mathbf{b}, \theta r\right) \subseteq \mathcal{M}$.

We also have that

$$\begin{aligned}
\max\{&\|\mathbf{b} - \mathbf{x}\|_2 \mid \mathbf{x} \in \mathrm{S}\left(\theta\mathbf{a}_0 + (1-\theta)\mathbf{b}, \theta r\right)\} \\
&= \max\{\|\mathbf{b} - (1-\theta)\mathbf{b} - \theta\mathbf{y}\|_2 \mid \mathbf{y} \in \mathrm{S}\left(\mathbf{a}_0, r\right)\} \\
&= \theta \max\{\|\mathbf{b} - \mathbf{y}\|_2 \mid \mathbf{y} \in \mathrm{S}\left(\mathbf{a}_0, r\right)\} \\
&\leq \theta \max\{\|\mathbf{z} - \mathbf{y}\|_2 \mid \mathbf{y}, \mathbf{z} \in \mathcal{M}\} \\
&\leq 2\theta R.
\end{aligned}$$

Now considering $\theta = \delta/(2R\|\mathbf{q}\|_2)$ and $\mathbf{c} = \theta\mathbf{a}_0 + (1-\theta)\mathbf{b}$, we get that $\theta r = t$ and $2\theta R = \delta/\|\mathbf{q}\|_2$. Therefore

$$\begin{aligned}
\mathrm{S}\left(\mathbf{c}, t\right) &\subseteq \mathcal{M} \cap \mathrm{S}\left(\mathbf{b}, \delta/\|\mathbf{q}\|_2\right) \\
&\subseteq \mathcal{M} \cap \{\mathbf{x} \in \mathbb{R}^n \mid |\langle \mathbf{q}, \mathbf{x}\rangle - \nu| \leq \delta\} \\
&= \{\mathbf{x} \in \mathcal{M} \mid \langle \mathbf{q}, \mathbf{x}\rangle \leq \nu + \delta\}.
\end{aligned}$$ □

We now present a sketch of the ellipsoid method in Algorithm 3.1.

Algorithm 3.1 Sketch of the ellipsoid method

Input: A vector $\mathbf{q} \in \mathbb{Q}^n$, a closed convex set $\mathcal{M} \subseteq \mathbb{R}^n$, with a valid quintuple $(\mathcal{M}; N, R, r, \mathbf{a}_0)$, and a scalar $\delta \in \mathbb{Q} \cap (0, 2R\|\mathbf{q}\|_2]$.
Output: $\widehat{\mathbf{x}} \in \mathcal{M}$ such that $\langle \mathbf{q}, \widehat{\mathbf{x}} \rangle \leq \min_{\mathbf{x} \in \mathbb{R}^n} \{\langle \mathbf{q}, \mathbf{x} \rangle \mid \mathbf{x} \in \mathcal{M}\} + \delta$.

```
let x_0 := 0,   E_0 := S(0, R)   and   x̂_0 := a_0.
let K := ⌈2N^2 ln(2R^2 ||q||_2 / (rδ))⌉.
for k = 0, ..., K − 1 do
   call a strong separation oracle for M and x_k.
   if the oracle asserts that x_k ∈ M then
      let h_k := −q
      if ⟨q, x_k⟩ ≤ ⟨q, x̂_k⟩ then
         let x̂_{k+1} := x_k.
      else
         let x̂_{k+1} := x̂_k.
      end if
   else
      let h_k equal the h given by the oracle.
      let x̂_{k+1} := x̂_k.
   end if
   let E_{k+1} be smallest ellipsoid containing {x ∈ E_k | ⟨h_k, x⟩ ≥ ⟨h_k, x_k⟩}.
   let x_{k+1} equal the center of E_{k+1}.
end for
output  x̂ := x̂_K.
```

We note that in Algorithm 3.1, for all $k = 0, \ldots, K$ we have $\widehat{\mathbf{x}}_k \in \mathcal{M}$ is the best feasible point found so far and

$$\{\mathbf{x} \in \mathcal{M} \mid \langle \mathbf{q}, \mathbf{x} \rangle \leq \langle \mathbf{q}, \widehat{\mathbf{x}}_k \rangle\} \subseteq \mathcal{E}_k.$$

From [GLS88], we have that for any $\mathbf{h} \in \mathbb{R}^n$ and ellipsoid $\mathcal{E} \subseteq \mathbb{R}^n$ with center $\mathbf{y}$, the smallest ellipsoid containing the set $\{\mathbf{x} \in \mathcal{E} \mid \langle \mathbf{h}, \mathbf{x} \rangle \geq \langle \mathbf{h}, \mathbf{y} \rangle\}$ has a volume less than or equal to $e^{-1/(2N)} \mathrm{Vol}(\mathcal{E})$. Therefore, for all k we have

$$\mathrm{Vol}\left(\{\mathbf{x} \in \mathcal{M} \mid \langle \mathbf{q}, \mathbf{x} \rangle \leq \langle \mathbf{q}, \widehat{\mathbf{x}}_k \rangle\}\right) \leq \mathrm{Vol}(\mathcal{E}_k) \leq e^{-k/(2N)} R^N \mathrm{Vol}\left(\mathrm{S}(\mathbf{0}, 1)\right).$$

From Lemma 3.12, we now see that for $K = \left\lceil 2N^2 \ln\left(2R^2\|\mathbf{q}\|_2/(r\delta)\right)\right\rceil$, we have that $\widehat{\mathbf{x}}_K \in \mathcal{M}$ and

$$\mathrm{Vol}\left(\{\mathbf{x} \in \mathcal{M} \mid \langle \mathbf{q}, \mathbf{x} \rangle \leq \langle \mathbf{q}, \widehat{\mathbf{x}}_k \rangle\}\right) \leq \mathrm{Vol}\left(\{\mathbf{x} \in \mathcal{M} \mid \langle \mathbf{q}, \mathbf{x} \rangle \leq \nu + \delta\}\right),$$

and thus $\langle \mathbf{q}, \widehat{\mathbf{x}}_K \rangle \leq \nu + \delta$ as required.

From this we then immediately get the following theorem.

Theorem 3.13. *Consider a* $\mathbf{q} \in \mathbb{Q}^n$*, a closed convex set* $\mathcal{M} \subseteq \mathbb{R}^n$*, with the quintuple* $(\mathcal{M}; N, R, r, \mathbf{a}_0)$ *being valid, and a* $\delta \in \mathbb{Q} \cap (0, 2R\|\mathbf{q}\|_2]$*. Furthermore, suppose that we have a strong separation oracle for* $\mathcal{M}$*. Then we can use the ellipsoid method, with at most* K *calls to the oracle, to find a* $\widehat{\mathbf{x}} \in \mathcal{M}$ *such that* $\langle \mathbf{q}, \widehat{\mathbf{x}} \rangle \leq \min_{\mathbf{x} \in \mathbb{R}^n} \{\langle \mathbf{q}, \mathbf{x} \rangle \mid \mathbf{x} \in \mathcal{M}\} + \delta$*, where* K *is polynomially bounded by the encoding lengths of* $\mathbf{q}$*, the quintuple and* δ*.*

In order to consider the polynomiality further, we need to worry about the encoding lengths of the variables throughout the algorithm, which necessitates rounding of the variables. This was done in [GLS88, Section 3.2] for a related problem, and by considering their proof we get the following theorem.

Theorem 3.14. *Consider a* $\mathbf{q} \in \mathbb{Q}^n$*, a closed convex set* $\mathcal{M} \subseteq \mathbb{R}^n$*, with the quintuple* $(\mathcal{M}; N, R, r, \mathbf{a}_0)$ *being valid, and a* $\delta \in \mathbb{Q} \cap (0, 2R\|\mathbf{q}\|_2]$*. Furthermore, suppose that we have a strong separation oracle for* $\mathcal{M}$*. Then we can use the ellipsoid method, with at most* K *calls to the oracle, to find a* $\widehat{\mathbf{x}} \in \mathcal{M}$ *such that* $\langle \mathbf{q}, \widehat{\mathbf{x}} \rangle \leq \min_{\mathbf{x} \in \mathbb{R}^n} \{\langle \mathbf{q}, \mathbf{x} \rangle \mid \mathbf{x} \in \mathcal{M}\} + \delta$*, where:*

i. K *is polynomially bounded by the encoding lengths of* $\mathbf{q}$*, the quintuple and* δ*,*

ii. the inputs to the oracle have encoding lengths which are polynomially bounded by the encoding lengths of $\mathbf{q}$*, the quintuple and* δ*,*

iii. the running time of the algorithm (excluding the running time of the oracle) is polynomially bounded by the encoding lengths of $\mathbf{q}$*, the quintuple and* δ*.*

In fact, this idea can be taken even further to show that we do not need a strong separation oracle, and in fact even a weak membership oracle will do. This was done in [GLS88, Chapter 4].

Theorem 3.15. *Consider a* $\mathbf{q} \in \mathbb{Q}^n$*, a closed convex set* $\mathcal{M} \subseteq \mathbb{R}^n$*, with the quintuple* $(\mathcal{M}; N, R, r, \mathbf{a}_0)$ *being valid, and a* $\delta \in \mathbb{Q}_{++}$ *such that we have* $\delta \leq \min\{r, 2R\|\mathbf{q}\|_2\}$*. Furthermore, suppose that we have a weak membership oracle for* $\mathcal{M}$*. Then we can use an extension of the ellipsoid method, with at most* K *calls to the oracle, in order to find an* $\widehat{\mathbf{x}} \in S(\mathcal{M}, \delta)$ *such that* $\langle \mathbf{q}, \widehat{\mathbf{x}} \rangle \leq \min_{\mathbf{x} \in \mathbb{R}^n} \{\langle \mathbf{q}, \mathbf{x} \rangle \mid \mathbf{x} \in S(\mathcal{M}, -\delta)\} + \delta$*, where:*

i. K *is polynomially bounded by the encoding lengths of* $\mathbf{q}$*, the quintuple and* δ*,*

ii. the inputs to the oracle have encoding lengths which are polynomially bounded by the encoding lengths of $\mathbf{q}$*, the quintuple and* δ*,* *iii. the running time of the algorithm (excluding the running time of the oracle) is polynomially bounded by the encoding lengths of* $\mathbf{q}$*, the quintuple and* δ*.*

The results from this section also extend trivially to the space of symmetric matrices.

3.3 Copositivity and Completely Positivity

In the paper [MK87] the following result was shown:

Theorem 3.16. *The strong membership problem for the copositive cone is a co-*$\mathcal{NP}$*-complete problem, i.e. both* $\mathcal{NP}$*-hard and in the class co-*$\mathcal{NP}$*.*

Due to this result, and the duality relationship between the copositive and completely positive cones, it has long been assumed that the strong membership problem for the completely positive cone is an $\mathcal{NP}$-complete problem, i.e. is both $\mathcal{NP}$-hard and in the class $\mathcal{NP}$. However, the technical details for this had not been considered until the papers [DG11, ABD12]. From these papers we get the following two theorems.

Theorem 3.17 ([DG11])**.** *We have that:*

i. both the strong and weak membership problems for the copositive cone are $\mathcal{NP}$*-hard,*

ii. both the strong and weak membership problems for the completely positive cone are $\mathcal{NP}$*-hard.*

Theorem 3.18 ([ABD12])**.** *The weak membership problem for the completely positive cone is in the class* $\mathcal{NP}$*.*

The proofs of both of these results are based on the ellipsoid method from the previous section. Theorem 3.17 was proven by using the ellipsoid method to consider the copositive and completely positive reformulations of the stable set problem from Section 1.3. Theorem 3.18 was proven using the ellipsoid method to consider a copositive optimisation problem which checks complete positivity of a matrix. In proving these results, many vital technical details must be considered. However, rather than going through all of the technical results from both papers in this thesis, we shall only consider those required for proving Theorem 3.18. This should then give the reader a flavour of the techniques that are required.

For a matrix $C \in \mathbb{Q}^{n\times n} \cap \mathcal{S}^n$, we consider the following copositive optimisation problem:

$$\begin{aligned} \min_{X\in\mathcal{S}^n} \quad & \langle C, X\rangle \\ \text{s.t.} \quad & X \in \mathcal{M}, \end{aligned} \tag{3.1}$$

$$\text{where } \mathcal{M} = \left\{X \in \mathcal{C}^n \;\middle|\; \tfrac{1}{2} \leq \langle I + \tfrac{1}{4n}E, X\rangle \leq \tfrac{3}{2}\right\}. \tag{3.2}$$

We then have the following theorem connecting this to the (weak) membership problem.

Theorem 3.19. *For $C \in \mathbb{Q}^{n\times n} \cap \mathcal{S}^n$ and $\varepsilon > 0$ we have that*

i. $\operatorname{Val}(3.1) \geq 0$ *if and only if* $C \in \mathcal{C}^{*n}$*,*

ii. $C \in S(\mathcal{C}^{*n}, -\varepsilon)$ *implies that* $\operatorname{Val}(3.1) \geq \varepsilon/(3n)$.

Proof. We shall prove each of these points separately.

i. In example 7.3 it shall be shown that $\left(I + \frac{1}{4n}E\right) \in \operatorname{int}\mathcal{C}^{*n}$. Combining this with Corollary 1.31 then gives the required result.

ii. First we note that $\|I + \frac{1}{4n}E\|_2 \leq \|(1 + \frac{1}{4n})E\|_2 = (1 + \frac{1}{4n})n < 3n/2$. We now note that if $C \in \mathrm{S}(\mathcal{C}^{*n}, -\varepsilon)$, then $\left(C - \frac{2}{3n}\varepsilon(I + \frac{1}{4n}E)\right) \in \mathcal{C}^{*n}$. Therefore, from the first part of this theorem, we get that

$$\begin{aligned} 0 &\leq \min_X \left\{\langle C - \tfrac{2}{3n}\varepsilon(I + \tfrac{1}{4n}E), X\rangle \;\middle|\; X \in \mathcal{M}\right\} \\ &= \min_X \left\{\langle C, X\rangle - \tfrac{2}{3n}\varepsilon\langle I + \tfrac{1}{4n}E, X\rangle \;\middle|\; \begin{array}{l} \tfrac{1}{2} \leq \langle I + \tfrac{1}{4n}E, X\rangle \leq \tfrac{3}{2}, \\ X \in \mathcal{C}^n \end{array}\right\} \\ &\leq \min_X \left\{\langle C, X\rangle - \tfrac{1}{3n}\varepsilon \;\middle|\; \tfrac{1}{2} \leq \langle I + \tfrac{1}{4n}E, X\rangle \leq \tfrac{3}{2}, \quad X \in \mathcal{C}^n\right\} \\ &= \operatorname{Val}(3.1) - \tfrac{1}{3n}\varepsilon. \end{aligned}$$

□

We next consider the following result on a valid quintuple for $\mathcal{M}$.

Theorem 3.20. *We have that the set $\mathcal{M}$ given in (3.2) is a closed convex set with a valid quintuple given by $(\mathcal{M}; N, R, r, A_0)$, where*

$$\begin{aligned} N &= \tfrac{1}{2}n(n+1), \\ R &= 5n, \\ r &= 1/(3n), \\ A_0 &= \tfrac{4}{4n+1}I. \end{aligned}$$

Proof. $\mathcal{M}$ is trivially a closed convex set and the value of N comes directly from the definition. We shall split the remainder of this proof into two parts:

i. Proof that $\mathcal{M} \subseteq \mathrm{S}(0, R)$:
We have $\mathcal{M} \subseteq \mathrm{S}(0, R)$ if and only if $\max_X\{\|X\|_2 \mid X \in \mathcal{M}\} \leq R$. We now let $\widehat{\mathcal{M}} = \{X \in \mathcal{C}^n \mid \langle I + \frac{1}{4n}E, X\rangle = \frac{3}{2}\}$. We have $\widehat{\mathcal{M}} \subseteq \mathcal{M}$ and thus $\max_X\{\|X\|_2 \mid X \in \mathcal{M}\} \geq \max_X\{\|X\|_2 \mid X \in \widehat{\mathcal{M}}\}$. Considering an arbitrary $X \in \mathcal{M}$, letting $\lambda = \frac{2}{3}\langle I + \frac{1}{4n}E, X\rangle \in [\frac{1}{3}, 1]$ and letting $Y = \lambda^{-1}X$, we have $Y \in \widehat{\mathcal{M}}$ and $\|Y\|_2 \geq \|X\|_2$. Therefore $\max_X\{\|X\|_2 \mid X \in \mathcal{M}\} \leq \max_X\{\|X\|_2 \mid X \in \widehat{\mathcal{M}}\}$, and thus

$$\mathcal{M} \subseteq \mathrm{S}(0, R) \qquad \Leftrightarrow \qquad \max_{X \in \mathcal{S}^n}\{\|X\|_2 \mid X \in \widehat{\mathcal{M}}\} \leq R.$$

We now consider the polyhedral cone

$$\mathcal{K} = \{X \in \mathcal{S}^n \mid (\boldsymbol{e}_i + \boldsymbol{e}_j)^{\mathsf{T}} X(\boldsymbol{e}_i + \boldsymbol{e}_j) \geq 0 \text{ for all } i, j\}.$$

It is trivial to see that $\mathcal{K}$ is a proper cone containing $\mathcal{C}$. From the fact that the number of linear inequalities describing this proper cone is equal to the dimension of the space containing it, each extreme ray of $\mathcal{K}$ is given by all but one of the inequalities being tight. From this, an alternative characterisation of $\mathcal{K}$ is $\operatorname{conic}(\mathcal{Y}_1 \cup \mathcal{Y}_2)$, where

$$\begin{aligned}\mathcal{Y}_1 &:= \{E_{ij} \mid 1 \leq i < j \leq n\},\\ \mathcal{Y}_2 &:= \{2E_{ii} - \textstyle\sum_{j \neq i} E_{ij} \mid i = 1, \ldots, n\},\end{aligned}$$

which together generate the extreme rays of $\mathcal{K}$. In fact an element E_{ij} of $\mathcal{Y}_1$ corresponds to all inequalities being set equal to zero except for the inequality $(\boldsymbol{e}_i + \boldsymbol{e}_j)^T X(\boldsymbol{e}_i + \boldsymbol{e}_j) \geq 0$, whilst an element $2E_{ii} - \sum_{j \neq i} E_{ij}$ of $\mathcal{Y}_2$ corresponds to all inequalities being set equal to zero except for the inequality $(2\boldsymbol{e}_i)^T X(2\boldsymbol{e}_i) \geq 0$.

It is straightforward to compute that

$$\begin{aligned}\langle I + \tfrac{1}{4n}E, E_{ij}\rangle &= \tfrac{1}{2n} && \text{for all } i \neq j,\\ \langle I + \tfrac{1}{4n}E, 2E_{ii} - \textstyle\sum_{j \neq i} E_{ij}\rangle &= \tfrac{3n+2}{2n} && \text{for all } i.\end{aligned}$$

We then get

$$\widehat{\mathcal{M}} \subseteq \widehat{\mathcal{K}} := \left\{X \in \mathcal{K} \;\middle|\; \left\langle I + \tfrac{1}{4n}E, X\right\rangle = \tfrac{3}{2}\right\} = \operatorname{conv}\left(3n\mathcal{Y}_1 \cup \tfrac{3n}{3n+2}\mathcal{Y}_2\right),$$

and therefore

$$\begin{aligned}\max\{\|X\|_2 \mid X \in \widehat{\mathcal{M}}\} &\leq \max\{\|X\|_2 \mid X \in \widehat{\mathcal{K}}\}\\ &= \max\{\|X\|_2 \mid X \in (3n\mathcal{Y}_1 \cup \tfrac{3n}{3n+2}\mathcal{Y}_2)\}\\ &= \max\left\{(3\sqrt{2})n, \tfrac{3n}{3n+2}\sqrt{2n+2}\right\}\\ &= (3\sqrt{2})n \leq R.\end{aligned}$$

ii. Proof that $\mathrm{S}(A_0, r) \subseteq \mathcal{M}$:
For A_0 as given above and $\lambda \in \mathbb{R}_+$, it can be seen from [HUS10] that $\mathrm{S}(A_0, \lambda) \subseteq \mathcal{C}$ if and only if $\lambda \leq \frac{4}{4n+1}$. A brief proof of this comes from noting that

$$\begin{aligned}
&\min_{\mathbf{v},X} \left\{ \mathbf{v}^{\mathsf{T}} X \mathbf{v} \;\middle|\; \begin{array}{l} \mathbf{v} \in \mathbb{R}^n_+, \ \|\mathbf{v}\|_2 = 1, \\ X \in \mathrm{S}(A_0, \lambda) \end{array} \right\} \\
&\qquad = \frac{4}{4n+1} + \lambda \min_{\mathbf{v},Z} \left\{ \mathbf{v}^{\mathsf{T}} Z \mathbf{v} \;\middle|\; \begin{array}{l} \mathbf{v} \in \mathbb{R}^n_+, \ \|\mathbf{v}\|_2 = 1, \\ Z \in \mathrm{S}(0, 1) \end{array} \right\},
\end{aligned}$$

$$\begin{aligned}
&\min_{\mathbf{v},Z} \left\{ \mathbf{v}^{\mathsf{T}} Z \mathbf{v} \;\middle|\; \begin{array}{l} \mathbf{v} \in \mathbb{R}^n_+, \ \|\mathbf{v}\|_2 = 1, \\ Z \in \mathrm{S}(0, 1) \end{array} \right\} \\
&\qquad \leq \min_{\mathbf{v}} \left\{ \mathbf{v}^{\mathsf{T}} \left(-\mathbf{v}\mathbf{v}^{\mathsf{T}} \right) \mathbf{v} \;\middle|\; \mathbf{v} \in \mathbb{R}^n_+, \ \|\mathbf{v}\|_2 = 1 \right\} = -1,
\end{aligned}$$

$$\begin{aligned}
&\min_{\mathbf{v},Z} \left\{ \mathbf{v}^{\mathsf{T}} Z \mathbf{v} \;\middle|\; \begin{array}{l} \mathbf{v} \in \mathbb{R}^n_+, \ \|\mathbf{v}\|_2 = 1, \\ Z \in \mathrm{S}(0, 1) \end{array} \right\} \\
&\qquad \geq \min_{\mathbf{v},Z} \left\{ -\|Z\|_2 \|\mathbf{v}\|_2^2 \;\middle|\; \begin{array}{l} \mathbf{v} \in \mathbb{R}^n_+, \ \|\mathbf{v}\|_2 = 1, \\ Z \in \mathrm{S}(0, 1) \end{array} \right\} = -1.
\end{aligned}$$

This implies that $\mathrm{S}(A_0, r) \subseteq \mathcal{C}$.

We now note that $\langle I + \frac{1}{4n} E, A_0 \rangle = 1$, and thus

$$\begin{aligned}
\max & \left\{ |\langle I + \tfrac{1}{4n} E, X \rangle - 1| \mid X \in \mathrm{S}(A_0, r) \right\} \\
&= r \max \left\{ \langle I + \tfrac{1}{4n} E, Y \rangle \mid Y \in \mathrm{S}(0, 1) \right\} \\
&= r \left\| I + \tfrac{1}{4n} E \right\|_2 \\
&\leq r \tfrac{3n}{2} = \tfrac{1}{2}.
\end{aligned}$$

□

The final technical result that we need before considering the ellipsoid method is the following.

Theorem 3.21. *We can construct a strong separation oracle for $\mathcal{M}$ given in* (3.2)*, such that for any $X \in \mathbb{Q}^{n \times n} \cap \mathcal{S}^n$, the oracle will either:*

i. assert that $X \in \mathcal{M}$, or

ii. assert that $\langle I + \frac{1}{4n} E, X \rangle < \frac{1}{2}$, or

iii. assert that $\langle I + \frac{1}{4n} E, X \rangle > \frac{3}{2}$, or

iv. find a $\mathbf{v} \in \mathbb{Q}^n_+$ such that $\mathbf{v}^{\mathsf{T}} X \mathbf{v} < 0$ and the encoding length of $\mathbf{v}$ is polynomially bounded in the encoding length of X.

Proof. Checking assertions (*ii*) and (*iii*) is a trivial task and from now on we shall assume that both of these assertions are false. We then have that assertion (*i*) is false if and only if $X \notin \mathcal{C}^n$. In [MK87] it was shown that for a matrix $X \in \mathbb{Q}^{n\times n} \cap \mathcal{S}^n \setminus \mathcal{C}^n$, there exists a $\mathbf{v} \in \mathbb{Q}^n_+$ such that $\mathbf{v}^\mathsf{T} X\mathbf{v} < 0$ and the encoding length of $\mathbf{v}$ is polynomially bounded in the encoding length of X. Therefore we have that assertion (*i*) is false if and only if there exists such a $\mathbf{v}$ as required in (*iv*). The problem of either asserting that such a $\mathbf{v}$ does not exist or finding such a $\mathbf{v}$, is equivalent to checking copositivity, and thus is an $\mathcal{NP}$-hard problem. However, as the encoding length of $\mathbf{v}$ is bounded, this can be done in finite time. $\square$

We now consider when $C \in \mathrm{S}\,(\mathcal{C}^{*n}, -\varepsilon)$ for $\varepsilon \in \mathbb{Q}_{++}$, and without loss of generality we may assume that $\varepsilon \leq 30n^2\|C\|_2$. From Theorem 3.19, we get that $\mathrm{Val}\,(3.1) \geq \varepsilon/(3n)$. We now use the analysis from the previous section on using the ellipsoid method for this optimisation problem, where

i. we let $\delta = \varepsilon/(3n)$,

ii. we use the quintuple given in Theorem 3.20,

iii. we use the strong separation oracle for $\mathcal{M}$ from Theorem 3.21.

From Theorem 3.14, we see that the ellipsoid method would then certify that $\mathrm{Val}\,(3.1) \geq 0$. We now consider $\mathbf{v}_1, \ldots, \mathbf{v}_L \in \mathbb{Q}^n_+$ being the vectors given when the separation oracle asserts statement (*iv*). We then get that the number of vectors, L, and the encoding lengths of the vectors, are polynomially bounded in the encoding lengths of C, the quintuple and δ, whose encoding lengths are in turn polynomially bounded in the encoding lengths of C and ε.

We now let

$$\widetilde{\mathcal{M}} = \left\{ X \in \mathcal{S}^n \;\middle|\; \begin{array}{l} \frac{1}{2} \leq \langle I + \frac{1}{4n}E, X\rangle \leq \frac{3}{2} \\ \mathbf{v}_i^\mathsf{T} X\mathbf{v}_i \geq 0 \text{ for all } i = 1,\ldots,L \\ (\boldsymbol{e}_i + \boldsymbol{e}_j)^\mathsf{T} X(\boldsymbol{e}_i + \boldsymbol{e}_j) \geq 0 \text{ for all } i,j = 1,\ldots,n \end{array} \right\}.$$

From the proof Theorem 3.20, we see that for the same values of N, R, r, A_0 we have that $(\widetilde{\mathcal{M}}; N, R, r, A_0)$ is a valid quintuple.

It is trivial to see that for any $X \in \mathcal{S}^n$ we have

$$X \in \mathcal{M} \qquad \Rightarrow \qquad X \in \widetilde{\mathcal{M}}.$$

Furthermore, for any X given in the algorithm to the strong separation oracle we have that either:

i. $X \in \mathcal{M}$, or

ii. $\langle I + \frac{1}{4n}E, X\rangle < \frac{1}{2}$, or

iii. $\langle I + \frac{1}{4n}E, X\rangle > \frac{3}{2}$, or

iv. there exists $i \in \{1, \dots, L\}$ such that $\mathbf{v}_i^\mathsf{T} X \mathbf{v}_i < 0$.

Therefore, for any X given in the algorithm to the strong separation oracle we have

$$X \in \mathcal{M} \qquad \Leftarrow \qquad X \in \widetilde{\mathcal{M}}.$$

This implies that the ellipsoid method was equivalently considering the following optimisation problem.

$$\begin{aligned} \min_{X \in \mathcal{S}^n} \quad & \langle C, X\rangle \\ \text{s.t.} \quad & X \in \widetilde{\mathcal{M}} \end{aligned}$$

Therefore the optimal value of this problem is greater than or equal to zero. We also have that A_0 is a strictly feasible point to this problem, which from considering the dual problem, along with Slater's condition, implies that the optimal value of the following problem is also greater than or equal to zero, as well as being attained:

$$\begin{aligned} \max_{\boldsymbol{\lambda}, \mathbf{z}, Z} \quad & \tfrac{1}{2}\lambda_1 - \tfrac{3}{2}\lambda_2 \\ \text{s.t.} \quad & C = (\lambda_1 - \lambda_2)(I + \tfrac{1}{4n}E) + \sum_{i \le j} (Z)_{ij} (\boldsymbol{e}_i + \boldsymbol{e}_j)(\boldsymbol{e}_i + \boldsymbol{e}_j)^\mathsf{T} + \sum_{i=1}^{L} z_i \mathbf{v}_i \mathbf{v}_i^\mathsf{T} \\ & \lambda_1, \lambda_2 \ge 0 \\ & (Z)_{ij}, z_i \ge 0 \qquad\qquad \text{for all } i, j. \end{aligned}$$

This optimisation problem could be used to generate at certificate that $C \in \mathcal{C}^*$, however a simpler certificate comes from noting that this optimisation problem having a nonnegative optimal value which is attained, implies that the following optimisation problem is feasible.

$$\begin{aligned} \min_{\mathbf{z}, Z} \quad & \sum_{i=1}^{L} z_i + \sum_{i \le j} (Z)_{ij} \\ \text{s.t.} \quad & C = \sum_{i \le j} (Z)_{ij} (\boldsymbol{e}_i + \boldsymbol{e}_j)(\boldsymbol{e}_i + \boldsymbol{e}_j)^\mathsf{T} + \sum_{i=1}^{L} z_i \mathbf{v}_i \mathbf{v}_i^\mathsf{T} \\ & (Z)_{ij}, z_i \ge 0 \qquad\qquad \text{for all } i, j. \end{aligned}$$

This final optimisation problem is a linear optimisation problem. We can then use the simplex method to find an optimal solution for this problem whose encoding length is polynomially bounded in the encoding length of the problem, which is in turn polynomially bounded in the encoding lengths of C and ε. This optimal solution is then a decomposition of C, which thus acts as a certificate for $C \in \mathcal{C}^{*n} \subseteq \mathrm{S}\left(\mathcal{C}^{*n}, \varepsilon\right)$.

In summary, we considered an arbitrary $C \in \mathbb{Q}^{n\times n} \cap \mathcal{S}^n$ and $\varepsilon \in \mathbb{Q}_{++}$, and supposed that $C \in \mathrm{S}\left(\mathcal{C}^{*n}, -\varepsilon\right)$. We then showed that there exists a certificate (in the form of a decomposition) which certifies that $C \in \mathcal{C}^{*n} \subseteq \mathrm{S}\left(\mathcal{C}^{*n}, \varepsilon\right)$ and whose encoding length is polynomially bounded in the encoding lengths of C and ε. Therefore, from the definition, we get that the weak membership problem for the completely positive cone is in the class $\mathcal{NP}$.

Chapter 4

Linear-time complete positivity detection and decomposition of sparse matrices*

As stated in the previous chapter, checking complete positivity and copositivity of a matrix are $\mathcal{NP}$-hard problems. However, in spite of the complexity of checking copositivity, for special cases there are efficient algorithms, even ones that run in linear-time. For example, in [Bom00] a method was discussed for checking whether a tridiagonal matrix is copositive in linear-time, while in [Ikr02] this was extended to acyclic matrices.

In this chapter, we will similarly consider special cases when we are able to check whether a matrix is completely positive in linear-time and, if so, find a minimal rank-one decomposition set for it.

Recall that a matrix X is completely positive if there exists a finite set $\mathcal{B} \subset \mathbb{R}^n_+$ such that

$$X = \sum_{\mathbf{b}\in\mathcal{B}} \mathbf{b}\mathbf{b}^\mathsf{T}.$$

In this case we say that $\mathcal{B}$ is a *rank-one decomposition set* of X.

One property that can be considered with regard to complete positivity is the *cp-rank*. The cp-rank of a completely positive matrix X is defined as

$$\text{cp-rank}(X) := \min\{|\mathcal{B}| \mid \mathcal{B} \text{ is a rank-one decomposition set of } X\}.$$

---

*Submitted as:

[DD12] P.J.C. Dickinson and M. Dür. Linear-time complete positivity detection and decomposition of sparse matrices. *SIAM Journal On Matrix Analysis and Applications*, 33(3):701–720, 2012

If X is a completely positive matrix, then from Carathéodory's theorem we have that cp-rank$(X) \leq \frac{1}{2}n(n+1)$, where n is the order of the matrix. This bound can be improved to cp-rank$(X) \leq \frac{1}{2}k(k+1)-1$, where $2 \leq k = \operatorname{rank} X \leq n$, as was shown in [BB03, LKF04]. It has also been conjectured in [DJL94] that

$$\text{cp-rank}(X) \leq \max\left\{n, \left\lfloor \frac{n^2}{4} \right\rfloor\right\},$$

and this has recently been proven to be correct for $n \leq 5$ [SMBJS12]. If a matrix is not completely positive, then its cp-rank is defined to be infinite [BR06].

We define a minimal rank-one decomposition set of a completely positive matrix X to be a rank-one decomposition set $\mathcal{B}$ such that $|\mathcal{B}| = \text{cp-rank}(X)$. In general this minimal rank-one decomposition set is not unique, as we shall see in Section 4.7. Properties of a rank-one decomposition of a completely positive matrix have been studied previously in [Dic10, GW80, Mar71], and in this chapter we will be investigating minimal rank-one decompositions for sparse matrices.

Related work

While the problem of finding a factorisation of a general completely positive matrix is still unsolved, the problem of factorising matrices with special structure has been studied before. Kaykobad [Kay87] proved that if a matrix is positive semidefinite and nonnegative and diagonally dominant, then it is completely positive. He also gives an easy procedure for constructing a factorisation. Berman and coauthors [BH87, BG88] considered matrices whose underlying graph has a special structure. In [BH87] they characterise completely positive matrices whose underlying graph is acyclic. They do not, however, use this characterisation for an algorithmic factorisation procedure. In [BG88] they study matrices with bipartite graphs and state a simple algorithmic procedure to factorise them.

The complete positivity of circular matrices has previously been studied in the papers [XL00, ZL00]. In [XL00] the authors characterise completely positive circular matrices of order greater than 3, but it seems unclear how this characterisation can actually be used to check algorithmically whether a circular graph is completely positive. In [ZL00] they give conditions for complete positivity of a circular matrix in terms of its comparison matrix. The proof of their result includes a method for finding a minimal rank-one decomposition set; however, this is a relatively complicated method and was not subjected to much analysis.

Li, Kummert, and Frommer [LKF04] show how — starting from an arbitrary factorisation of a matrix $X \in \mathcal{S}^n$, with $n > 1$ — one can obtain a smaller factorisation $X = \sum_{\mathbf{b}\in\mathcal{B}} \mathbf{b}\mathbf{b}^\mathsf{T}$ with $|\mathcal{B}| = \frac{1}{2}n(n+1) - 1$.

Shaked-Monderer [SM09] considers matrices which are positive semidefinite and nonnegative with rank r and have an $r \times r$ principal submatrix that is diagonal. This corresponds to the graph of the matrix having a maximal stable set of size r. Such a matrix is shown to be completely positive, and a factorisation is immediate from the proof. Kalofolias and Gallopoulos [KG12] extend this result and construct a factorisation of completely positive rank-two matrices.

Finally, Dong, Lin, and Chu [DLC12] provide a heuristic method for the so-called (nonsymmetric) nonnegative rank factorisation, i.e. finding a decomposition $X = UV$ of X with U, V nonnegative but not necessarily $U = V^T$ (which would correspond to our setting). Their procedure can be applied to completely positive matrices and would be able to heuristically check whether cp-rank$(X) = \operatorname{rank}(X)$ and, if affirmative, compute a factorisation of X.

This chapter will provide a unified approach to these ideas and extend the domain of cases where a factorisation can be found. We will present an algorithmic method for this and pay special attention to the run-time of this algorithm. One of our results will be that — as in the copositive case studied in [Bom00, Ikr02] — for tridiagonal and acyclic matrices, complete positivity can be checked in linear-time. Our method could also be used for preprocessing a matrix which we wish to test for complete positivity in order to reduce the problem.

Graphs of a Matrix

For a matrix $A \in \mathcal{S}^n$, we define $G(A)$ to be the *underlying graph* of A such that $G(A) = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \{1, \dots, n\}$ and $\mathcal{E} = \{(ij) \mid i < j,\ (A)_{ij} \neq 0\}$. When we talk of an index i of A having a certain degree, we are referring to the degree of the vertex i in the graph $G(A)$ having this degree. Similarly, when we refer to graph properties of a matrix A, for example being acyclic, circular or connected, we are referring to the properties of the graph $G(A)$. In this thesis, when we refer to a circular graph, we mean a graph only consisting of a single cycle. We will use the phrase *component submatrix* for a principal submatrix whose graph is a connected component in the graph of the full matrix. Finally a weighted-graph of A refers to $G(A)$ with weights on the vertices and edges equal to the corresponding values in A. We use this in order to be able to consider certain structures in a matrix with more ease.

4.1 Rank-One Decomposition

We will now look at some basic properties of (minimal) rank-one decomposition sets of sparse completely positive matrices.

As previously noted in Section 1.1, we have that a completely positive matrix is nonnegative and, if an on-diagonal element of a completely positive matrix is equal to zero, then all the off-diagonal elements on this row and column are also equal to zero. We can in fact check whether these necessary conditions hold in linear-time.

We now look at how the graph of a completely positive matrix corresponds to the support of the vectors in a rank-one decomposition of the matrix. We consider a completely positive matrix $X \neq 0$ with a rank-one decomposition set $\mathcal{B}$. For a vector $\mathbf{b} \in \mathcal{B}$ we have that the set $\{i \mid (\mathbf{b})_i > 0\}$ is a clique of $G(X)$. Correspondingly, if a set of vertices $\mathcal{J} \subseteq \{1, \ldots, n\}$ is not a clique of $G(X)$, then there cannot be a vector $\mathbf{b} \in \mathcal{B}$ such that $\{i \mid (\mathbf{b})_i > 0\} = \mathcal{J}$. Therefore we need only consider each component submatrix of a matrix separately, and it should be noted that using for example a breadth-first search, we can split a matrix into its component submatrices in linear-time.

From now on, without loss of generality, we shall assume that the matrices that we wish to analyse are nonnegative and connected and have all on-diagonal elements strictly positive.

We finish this section by looking at a special property which always holds for at least one minimal rank-one decomposition of a completely positive matrix.

Theorem 4.1. *For any completely positive matrix A there exists a minimal rank-one decomposition of it such that no two vectors in the decomposition have the same support.*

Proof. Consider two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n_{++}$. We define the following:

$$\begin{aligned}
\mu &= \max\{\lambda \in \mathbb{R} \mid \mathbf{b} - \lambda\mathbf{a} \geq 0\}, \\
\mathbf{c} &= \frac{1}{\sqrt{1+\mu^2}}(\mathbf{b} - \mu\mathbf{a}), \\
\mathbf{d} &= \frac{1}{\sqrt{1+\mu^2}}(\mathbf{a} + \mu\mathbf{b}).
\end{aligned}$$

Then we have that

$$\begin{aligned}
&\mathbf{c} \in \mathbb{R}^n_+ \setminus \mathbb{R}^n_{++}, \\
&\mathbf{d} \in \mathbb{R}^n_{++}, \\
&\mathbf{a}\mathbf{a}^\mathsf{T} + \mathbf{b}\mathbf{b}^\mathsf{T} = \mathbf{c}\mathbf{c}^\mathsf{T} + \mathbf{d}\mathbf{d}^\mathsf{T}.
\end{aligned}$$

This can easily be extended to any two vectors with the same support. We can now take any minimal rank-one decomposition of a completely positive matrix and use this method to get the desired property. □

4.2 Indices of Degree Zero or One

In this section, we will look at how we can reduce the problem of checking whether a matrix is completely positive by considering indices of the matrix with degree zero or one. Recall that we have defined the degree of an index to be the degree of the corresponding vertex in the graph of the matrix.

Degree-zero indices are themselves component submatrices and so can be considered separately. As they are size 1×1 matrices, checking them for complete positivity and, if this is found, providing a minimal rank-one decomposition set are trivial tasks.

In order to see how to deal with indices of a higher degree, we first consider the following theorem.

Theorem 4.2. *We define the matrices $X, Y_\theta, Z_\theta \in \mathcal{S}^n$ as*

$$X = \begin{pmatrix} A_1 & \mathbf{a}_1 & 0 \\ \mathbf{a}_1^\mathsf{T} & \alpha & \mathbf{a}_2^\mathsf{T} \\ 0 & \mathbf{a}_2 & A_2 \end{pmatrix}, \qquad Y_\theta = \begin{pmatrix} A_1 & \mathbf{a}_1 & 0 \\ \mathbf{a}_1^\mathsf{T} & \theta & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad Z_\theta = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \alpha - \theta & \mathbf{a}_2^\mathsf{T} \\ 0 & \mathbf{a}_2 & A_2 \end{pmatrix},$$

where $\alpha, \theta \in \mathbb{R}$, $A_1 \in \mathcal{S}^p$, $A_2 \in \mathcal{S}^q$, $\mathbf{a}_1 \in \mathbb{R}^p$, $\mathbf{a}_2 \in \mathbb{R}^q$, $p, q, n \in \mathbb{Z}_{++}$ and $n = p + q + 1$. Then the following three statements are equivalent:

i. X is completely positive.

ii. There exists θ such that Y_θ and Z_θ are completely positive.

iii. $\varphi := \min\{\theta \mid Y_\theta \in \mathcal{C}^*\}$ is finite and Z_φ is completely positive.

Proof. We first note that the value of the minimisation in (*iii*) is either infinity or it is attained (in which case Y_φ is completely positive). It can now be immediately seen that (*iii*) $\Rightarrow$ (*ii*) $\Rightarrow$ (*i*). From considering the cliques of $G(X)$ we see that (*i*) $\Rightarrow$ (*ii*). It is also a simple task to show that (*ii*) $\Rightarrow$ (*iii*), by noting that if $Z_\theta \in \mathcal{C}^*$ for some θ, then $Z_\phi \in \mathcal{C}^*$ for all $\phi \leq \theta$, which completes the proof. □

From this theorem we see that in some special cases, when it is relatively easy to find $\varphi = \min\{\theta \mid Y_\theta \in \mathcal{C}^*\}$, we can reduce the problem of checking whether X is completely positive to checking whether the smaller nonzero principal submatrix of Z_φ is completely positive. An example of when it is relatively easy to find φ is when the underlying graph given by Y_θ is a *completely positive graph.* A completely positive graph is defined to be a graph such that for all Y with this underlying graph we have that $Y \in \mathcal{C}^*$ if and only if $Y \in \mathcal{S}_+ \cap \mathcal{N}$. A characterisation of these graphs is that they have no odd cycles of length greater than or equal to five [KB92]. This means that in such a case we have $\varphi = \min\{\theta \mid Y_\theta \in \mathcal{S}_+ \cap \mathcal{N}\}$, and this optimisation problem can be solved in polynomial time up to any required accuracy [NN93].

In the following theorem, we now look at a very simple but very useful special case, which was first considered by Berman and Hershkowitz [BH87]. They took a different approach to this problem and proved part (i) and a similar result to part (iv), except that they gave an inequality relation, whereas we will give an equality.

Theorem 4.3. *Let $X \in \mathcal{N}^n$, $Y \in \mathcal{S}^n$ be given as,*

$$X = \begin{pmatrix} \alpha & \beta & 0 \\ \beta & \gamma & \mathbf{a}^\mathsf{T} \\ 0 & \mathbf{a} & A \end{pmatrix}, \qquad Y = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \gamma - \frac{1}{\alpha}\beta^2 & \mathbf{a}^\mathsf{T} \\ 0 & \mathbf{a} & A \end{pmatrix},$$

where $\alpha, \beta, \gamma \in \mathbb{R}_+$, $\alpha \neq 0$, $\mathbf{a} \in \mathbb{R}^{n-2}_+$, and $A \in \mathcal{N}^{n-2}$. Then we have the following

i. $X \in \mathcal{C}^* \;\Leftrightarrow\; Y \in \mathcal{C}^*$.

ii. For $X \in \mathcal{C}^*$, if we let $\mathcal{B}_Y \subset \mathbb{R}^n_+$ be a rank-one decomposition set of Y, then the following set is a rank-one decomposition set of X:

$$\mathcal{B}_X = \mathcal{B}_Y \cup \left\{ \begin{pmatrix} \sqrt{\alpha} \\ \beta/\sqrt{\alpha} \\ \mathbf{0} \end{pmatrix} \right\}.$$

iii. If in (ii) $\mathcal{B}_Y$ is a minimal rank-one decomposition set of Y, then $\mathcal{B}_X$ is a minimal rank-one decomposition set of X.

iv. We have that $\text{cp-rank}(X) = \text{cp-rank}(Y) + 1$ *(where $\infty + 1 := \infty$).*

Proof. From [MM62] we have that $\mathcal{C}^{*2} = \mathcal{S}^2_+ \cap \mathcal{N}^2$. Therefore

$$\begin{pmatrix} \alpha & \beta \\ \beta & \theta \end{pmatrix} \in \mathcal{C}^* \Leftrightarrow \theta \geq \beta^2/\alpha,$$

and using this, Theorem 4.2 gives us a proof for (i). Part (ii) is trivial to prove and part (iv) comes directly from part (iii). We will now prove part (iii). Another way of expressing part (iii) is that there exists a minimal rank-one decomposition of X given by $\mathcal{B}_X$ such that

$$\{v \in \mathcal{B}_X \mid (v)_1 > 0\} = \left\{ \begin{pmatrix} \sqrt{\alpha} \\ \beta/\sqrt{\alpha} \\ \mathbf{0} \end{pmatrix} \right\}. \tag{4.1}$$

Due to Theorem 4.1, there exists a minimal rank-one decomposition of X given by $\widehat{\mathcal{B}}_X$ such that no two vectors in the decomposition have the same support. If property (4.1) holds, then we are done. If not, then by considering the cliques of $G(X)$, we see that there exists $\varphi \in \mathbb{R}$ such that $0 < \varphi < \sqrt{\alpha}$ and

$$\left\{ v \in \widehat{\mathcal{B}}_X \;\middle|\; (v)_1 > 0 \right\} = \left\{ \begin{pmatrix} \varphi \\ \beta/\varphi \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sqrt{\alpha - \varphi^2} \\ 0 \\ \mathbf{0} \end{pmatrix} \right\}.$$

It is trivial to see that

$$\begin{aligned} &\begin{pmatrix} \varphi \\ \beta/\varphi \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \varphi \\ \beta/\varphi \\ \mathbf{0} \end{pmatrix}^{\mathsf{T}} + \begin{pmatrix} \sqrt{\alpha - \varphi^2} \\ 0 \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \sqrt{\alpha - \varphi^2} \\ 0 \\ \mathbf{0} \end{pmatrix}^{\mathsf{T}} \\ &\qquad = \begin{pmatrix} \sqrt{\alpha} \\ \beta/\sqrt{\alpha} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \sqrt{\alpha} \\ \beta/\sqrt{\alpha} \\ \mathbf{0} \end{pmatrix}^{\mathsf{T}} + \begin{pmatrix} 0 \\ \eta \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} 0 \\ \eta \\ \mathbf{0} \end{pmatrix}^{\mathsf{T}} \end{aligned}$$

where $\eta = \beta\sqrt{(\alpha - \varphi^2)/(\alpha\varphi^2)}$.

We can use this fact to obtain an alternative minimal rank-one decomposition of X such that property (4.1) does hold. □

From this theorem and the result on degree zero indices, we can now construct Algorithm 4.1 for reducing the problem of checking whether a matrix is completely positive and, if so, finding a (minimal) rank-one decomposition set. Although Berman and Hershkowitz [BH87] considered a similar method of going through a matrix, this was only to prove that an acyclic matrix is completely positive if and only if it is both nonnegative and positive semidefinite. They did not consider how it could also be used to produce a rank-one decomposition set or its computation time. We now give the following results for this algorithm.

Theorem 4.4. *Algorithm 4.1 gives the required output, and, if it does not produce the message "$X \notin \mathcal{C}^*$", then the X' and $\mathcal{B}$ produced will have the following properties:*

i. $X = X' + \sum_{\mathbf{b}\in\mathcal{B}} \mathbf{b}\mathbf{b}^{\mathsf{T}}$.

ii. $\text{cp-rank}(X) = \text{cp-rank}(X') + |\mathcal{B}|$.

iii. If an index i of X' has degree zero then $(X')_{ii} = 0$.

iv. X' has no indices of degree one.

Algorithm 4.1 Reducing the problem of checking for complete positivity.

Input: A matrix $X \in \mathcal{N}^n$ such that $(X)_{ii} > 0$ for all $i = 1, \dots, n$.
Output: Either "$X \notin \mathcal{C}^*$" or a matrix $X' \in \mathcal{S}^n$ and finite set $\mathcal{B} \subset \mathbb{R}^n_+$ (see Theorem 4.4).

initiate a set $\mathcal{B} = \emptyset$.
initiate a set $\mathcal{R} = \{1, \dots, n\}$ to keep track of indices remaining.
analyse X producing
    i. a set $\mathcal{J} \subseteq \mathcal{R}$ of indices with degree zero or one,
    ii. a vector $\mathbf{d} \in \mathbb{Z}^n$ such that $(\mathbf{d})_i$ is defined to be degree of index i,
    iii. a set $\{\mathfrak{N}_1, \dots, \mathfrak{N}_n\}$ s.t. $\mathfrak{N}_i := \{j \mid j \text{ is neighbour of } i \text{ in } G(X)\}$.
while $\mathcal{J} \neq \emptyset$ **do**
    pick an $i \in \mathcal{J}$ to analyse.
    update $\mathcal{R} \leftarrow \mathcal{R} \setminus \{i\}$, $\quad \mathcal{J} \leftarrow \mathcal{J} \setminus \{i\}$
    if $(\mathbf{d})_i = 0$ **then**
        update $\mathcal{B} \leftarrow \mathcal{B} \cup \{\sqrt{(X)_{ii}}\, \boldsymbol{e}_i\}$
        update $(X)_{ii} \leftarrow 0$
    else
        find $j \in \mathfrak{N}_i \cap \mathcal{R}$
        update $(X)_{jj} \leftarrow (X)_{jj} - (X)_{ij}^2 / (X)_{ii}$
        if $(X)_{jj} < 0$ **or** $\big((X)_{jj} = 0$ **and** $(\mathbf{d})_j \geq 2\big)$ **then**
            output "$X \notin \mathcal{C}^*$"
            exit
        end if
        update $\mathcal{B} \leftarrow \mathcal{B} \cup \left\{ \sqrt{(X)_{ii}}\, \boldsymbol{e}_i + \left((X)_{ij} / \sqrt{(X)_{ii}}\right) \boldsymbol{e}_j \right\}$
        update $(X)_{ij} \leftarrow 0$, $\quad (X)_{ii} \leftarrow 0$, $\quad (\mathbf{d})_j \leftarrow (\mathbf{d})_j - 1$
        if $(\mathbf{d})_j = 1$ **then**
            update $\mathcal{J} \leftarrow \mathcal{J} \cup \{j\}$
        else if $(\mathbf{d})_j = 0$ **and** $(X)_{jj} = 0$ **then**
            update $\mathcal{J} \leftarrow \mathcal{J} \setminus \{j\}$, $\quad \mathcal{R} \leftarrow \mathcal{R} \setminus \{j\}$
        end if
    end if
end while
output $\mathcal{B}$
output $X' \leftarrow X$

Also, provided that our inputs and outputs of the matrices and vectors in Algorithm 4.1 were "efficient" (see proof), then this algorithm runs in linear-time.

Proof. From going through the algorithm and using Theorem 4.3, this is trivial to prove. It should be noted that we required the matrices and vectors to be inputted/outputted efficiently. Firstly, this means dealing with the square roots. Secondly, this is because the inputting/outputting of a full vector or matrix would involve respectively n and n^2 entries, which would limit the algorithm to working in quadratic time. However, a more efficient way of specifying a sparse vector or matrix is to give only its nonzero entries, and we are required to do this in order for the algorithm to work in linear-time. $\square$

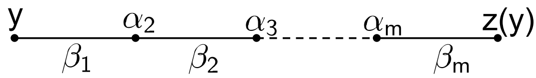
For Algorithm 4.1 we can see that if X'' is the maximal principal submatrix of X' such that no row/column is equal to zero, then $G(X'')$ is the maximal induced subgraph of $G(X)$ such that $G(X'')$ has no vertices of degree zero or one. This means that if X was acyclic, for example, tridiagonal, then in linear-time either the algorithm would output $X \notin \mathcal{C}^*$ or we would have $X' = 0$ and therefore a certificate of complete positivity in the form of a minimal rank-one decomposition set $\mathcal{B}$. We can also see that such a minimal rank-one decomposition set would be of cardinality $n-1$ or n. It was found in [BSM03, Theorem 3.7] that this number is actually equal to the rank of the matrix.

It should also be noted that the choice of the next vertex to consider in step 5 of Algorithm 4.1 affects the way in which the algorithm goes through the vertices and can lead to a different set $\mathcal{B}$ at the end of the algorithm. If we simply go through the vertices in $\mathcal{J}$ in numerical order, then given a permutation matrix P the algorithm will not necessarily return the same solution (up to permutation) when X and PXP^T are inputted.

4.3 Chains

We define a chain of a graph to be a simple path in it such that all vertices within the path (excluding the two end vertices) have degree equal to two. This is equivalent to this part of the matrix being tridiagonal. The form of chain that we shall consider is shown in Fig. 4.1, where we let the internal vertices and the edges of the chain have fixed weightings (given by the α's and β's respectively), we let y be a variable, and we let $z(y)$ be the minimum allowable value such that the chain is completely positive. We then consider how Algorithm 4.1 would run through the chain starting at y, moving through the vertices in turn and being used to give us the value of $z(y)$ for each y. We show that rather than having to recompute the algorithm for different values of y, we can instead find a simple formula linking y and $z(y)$. We consider this, not only to help improve our understanding of Algorithm 4.1, but also due to the useful applications that this method will provide in Sections 4.4 and 4.5.

Figure 4.1: A chain we consider the algorithm working through with $y > 0$ and $\alpha_i, \beta_i > 0$ for all i.



For $i \geq 2$ the algorithm would take each vertex from α_i to $f_i(y)$ and then to zero, where $f_1(y) := y$ and $f_i(y) := \alpha_i - (\beta_{i-1}^2 / f_{i-1}(y))$ for $i = 2, \ldots, m$.

We require that $f_i(y) > 0$ for all i, and we also have $z(y) = \beta_m^2 / f_m(y)$.

We now present the following lemmas, which will be used to remove the recursion from y, thus meaning that we do not need to consider the recursion separately for each different value of y.

Lemma 4.5. *We have that*

$$f_i(y) = \frac{\lambda_i y - \mu_i}{\lambda_{i-1} y - \mu_{i-1}} \qquad \text{for } i = 1, \ldots, m,$$

where $\lambda_0 = 0$, $\mu_0 = -1$, $\lambda_1 = 1$, $\mu_1 = 0$,

$$\lambda_i = \alpha_i \lambda_{i-1} - \beta_{i-1}^2 \lambda_{i-2} \qquad \text{for } i = 2, \ldots, m,$$
$$\mu_i = \alpha_i \mu_{i-1} - \beta_{i-1}^2 \mu_{i-2} \qquad \text{for } i = 2, \ldots, m.$$

Also we have that the requirement "$f_i(y) > 0$ for all i" is equivalent to

$$\lambda_i y > \mu_i \qquad \text{for all } i. \tag{4.2}$$

Proof. This is trivial to prove by induction. □

Lemma 4.6. *We always get that* $\mu_i \lambda_{i+1} < \mu_{i+1} \lambda_i$ *for all* $i = 0, \ldots, m-1$.

Proof. We have that $\mu_0 \lambda_1 = -1 < 0 = \mu_1 \lambda_0$, and for $i \geq 1$,

$$\begin{aligned} \mu_{i+1}\lambda_i - \mu_i \lambda_{i+1} &= (\alpha_{i+1}\mu_i - \beta_i^2 \mu_{i-1})\lambda_i - \mu_i(\alpha_{i+1}\lambda_i - \beta_i^2 \lambda_{i-1}) \\ &= \beta_i^2 (\mu_i \lambda_{i-1} - \mu_{i-1}\lambda_i). \end{aligned}$$

We can now use proof by induction. □

Lemma 4.7. *Requirement* (4.2) *implies that* $\lambda_i, \mu_i > 0$ *for all* $i \geq 2$.

Proof. We shall again use proof by induction. For the case when $i = 2$ we have that $\lambda_2 = \alpha_2 > 0$ and $\mu_2 = \beta_1^2 > 0$. Therefore the statement is true for $i = 2$. Now, for the sake of induction, suppose that it is true for $i = k - 1$. From Lemma 4.6 and requirement (4.2) we get that

$$\frac{\mu_{k-1}}{\lambda_{k-1}} < y \qquad \text{and} \qquad \frac{\mu_{k-1}\lambda_k}{\lambda_{k-1}} < \mu_k < \lambda_k y.$$

From this we see that we have $\lambda_k, \mu_k > 0$. □

Lemma 4.8. *If $\lambda_i > 0$ for all $i \geq 1$, then* $\max\left\{\frac{\mu_i}{\lambda_i} \;\middle|\; i = 1, \ldots, m\right\} = \frac{\mu_m}{\lambda_m}$.

Proof. This is simple to prove using Lemma 4.6. □

Theorem 4.9. *Requirement* (4.2) *is equivalent to*

$$y > \frac{\mu_m}{\lambda_m} \qquad \textit{and} \qquad \lambda_i > 0 \quad \textit{for all } i \geq 2.$$

Proof. The proof comes trivially from Lemmas 4.7 and 4.8. □

Method 4.10. Therefore the problem of going through the chain can be split into the following three parts:

i. Compute the following:

$$\begin{aligned}
&\lambda_0 = 0,\ \mu_0 = -1,\ \lambda_1 = 1,\ \mu_1 = 0, \\
&\lambda_i = \alpha_i\lambda_{i-1} - \beta_{i-1}^2\lambda_{i-2} \qquad \text{for } i = 2, \ldots, m, \\
&\mu_i = \alpha_i\mu_{i-1} - \beta_{i-1}^2\mu_{i-2} \qquad \text{for } i = 2, \ldots, m.
\end{aligned}$$

ii. Check that $\lambda_i > 0$ for all $i \geq 2$; otherwise it cannot be part of a completely positive matrix.

iii. Require that $y > \mu_m/\lambda_m$ and

$$z(y) = \frac{\beta_m^2(\lambda_{m-1}y - \mu_{m-1})}{\lambda_m y - \mu_m}.$$

In the following two sections we look at two alternative ways in which this result can be used.

4.4 Matrices with circular graphs

If Algorithm 4.1 did not determine whether the original matrix was completely positive or not, then the degree of the indices in the remaining matrix is strictly greater than one, and so the simplest form that it can take is being a circular matrix, where we recall that a circular matrix is one with an underlying circular graph and a circular graph is a graph only consisting of a single cycle. This is also sometimes referred to as a cycle graph.

The complete positivity of circular matrices has previously been studied in the papers [XL00, ZL00]. In [XL00], the authors found a necessary and sufficient condition for a circular matrix (of order greater than 3) to be completely positive; however it is unclear how this result can actually be used to check whether a circular graph is completely positive. In [ZL00] they showed that a circular matrix (of order greater than 3) is completely positive if and only if the determinant of its comparison matrix is nonnegative. The paper also included a method for finding a minimal rank-one decomposition set of a circular matrix (of order greater than 3); however, this was included only for the purpose of providing a proof to a theorem related to the number of minimal rank-one decompositions that a circular matrix has. As a result, this method was not subjected to much analysis and is relatively complicated.

In this section, we will use the results from Section 4.3 to develop an alternative algorithm for checking whether a circular matrix is complete positive and, if so, providing a minimal rank-one decomposition set of the matrix. It will also be seen that this method runs in linear-time.

We will begin by considering the following two theorems.

Theorem 4.11 ([BSM03, Remark 3.3])**.** *If A is a triangle-free, connected completely positive matrix which is not acyclic, then the cp-rank of A is equal to the number of edges in the graph $G(A)$.*

Theorem 4.12 ([BSM03, Theorem 3.2])**.** *Let $A \in \mathcal{C}^{*n}$, with $n \leq 3$. Then* $\text{cp-rank}(A) = \operatorname{rank} A$.

As we see from these theorems, we should consider the cases of $n = 3$ and $n > 3$ separately. We will first extend our method from Section 4.3 for the case when $n > 3$. We let $A \in \mathcal{N}^n$ be a circular matrix and without loss of generality suppose that

$$A = \begin{pmatrix} \alpha_1 & \beta_1 & 0 & \cdots & 0 & \beta_n \\ \beta_1 & \alpha_2 & \beta_2 & \cdots & 0 & 0 \\ 0 & \beta_2 & \alpha_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha_{n-1} & \beta_{n-1} \\ \beta_n & 0 & 0 & \cdots & \beta_{n-1} & \alpha_n \end{pmatrix}. \tag{4.3}$$

If A is completely positive, then due to its cp-rank being equal to n (Theorem 4.11) and by considering its cliques, we see that its minimal decompositions are of the form

$$\mathcal{B} = \left\{ \begin{pmatrix} \upsilon_1 \\ \omega_1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \upsilon_2 \\ \omega_2 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ \upsilon_{n-1} \\ \omega_{n-1} \end{pmatrix}, \begin{pmatrix} \omega_n \\ 0 \\ 0 \\ \vdots \\ 0 \\ \upsilon_n \end{pmatrix} \right\}.$$

From this it can be seen that A is completely positive if and only if there exists a $y = \upsilon_1^2$ such that the chain in Fig. 4.1 is completely positive with $z(y) = \alpha_1 - y$, and it can easily be seen how minimal rank-one decompositions of the original matrix A and the chain are related. From Section 4.3 we now get the following linear-time method for analysing the matrix.

Method 4.13. The problem of determining whether A given in (4.3) is completely positive is equivalent to computing

$$\begin{aligned}
&\lambda_0 = 0,\ \mu_0 = -1,\ \lambda_1 = 1,\ \mu_1 = 0, \\
&\lambda_i = \alpha_i \lambda_{i-1} - \beta_{i-1}^2 \lambda_{i-2} \qquad \text{for } i = 2, \ldots, n, \\
&\mu_i = \alpha_i \mu_{i-1} - \beta_{i-1}^2 \mu_{i-2} \qquad \text{for } i = 2, \ldots, n,
\end{aligned}$$

checking that $\lambda_i > 0$ for all $i \geq 2$ and solving

$$\begin{aligned}
\text{find} \quad & y \\
\text{s.t.} \quad & y > \mu_n / \lambda_n \\
& 0 = \lambda_n y^2 + (\beta_n^2 \lambda_{n-1} - \alpha_1 \lambda_n - \mu_n) y + (\alpha_1 \mu_n - \beta_n^2 \mu_{n-1}).
\end{aligned}$$

We also have that if y is a feasible solution, then the corresponding minimal rank-one decomposition is

$$\mathcal{B} = \left\{ \begin{pmatrix} \upsilon_1 \\ \omega_1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \upsilon_2 \\ \omega_2 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ \upsilon_{n-1} \\ \omega_{n-1} \end{pmatrix}, \begin{pmatrix} \omega_n \\ 0 \\ 0 \\ \vdots \\ 0 \\ \upsilon_n \end{pmatrix} \right\},$$

$$\begin{aligned}
\text{where } \upsilon_i &= \sqrt{\frac{\lambda_i y - \mu_i}{\lambda_{i-1} y - \mu_{i-1}}}, \\
\omega_i &= \beta_i / \upsilon_i.
\end{aligned} \qquad \text{for all } i$$

We note that this method even checks the complete positivity in linear-time of circular matrices such that their order is odd and greater than or equal to 5, even though for these types of matrices, their underlying graph is not a completely positive graph.

For completeness we will now also consider how to test whether a strictly positive matrix $X \in \mathcal{S}^3$ is completely positive and, if so, find a minimal rank-one decomposition set for it. In order to do this we need the following lemmas, in which we will consider the matrix

$$X = \begin{pmatrix} \alpha_1 & \beta_1 & \beta_3 \\ \beta_1 & \alpha_2 & \beta_2 \\ \beta_3 & \beta_2 & \alpha_3 \end{pmatrix}, \qquad \text{where } \alpha, \beta \in \mathbb{R}^3_{++} \text{ and } \alpha_3\beta_1^2 \leq \alpha_1\beta_2^2. \tag{4.4}$$

It should be noted that we can always permute a 3×3 strictly positive symmetric matrix so that the required inequalities hold.

Lemma 4.14. *For X given in* (4.4) *we have $X \in \mathcal{C}^*$ if and only if $\beta_1^2 \leq \alpha_1\alpha_2$, $\beta_2^2 \leq \alpha_2\alpha_3$, $\beta_3^2 \leq \alpha_3\alpha_1$, and* $\det(X) \geq 0$.

Proof. From [MM62] we have that $\mathcal{C}^{*3} = \mathcal{S}^3_+ \cap \mathcal{N}^3$. From the conditions in (4.4) we have that $X \in \mathcal{N}^3$. It is known that a matrix is positive semidefinite if and only if all its principal minors are nonnegative [Mur03, page 40]. This combined with the fact that the diagonal entries of X are nonnegative gives us the required result. □

Lemma 4.15. *For $X \in \mathcal{C}^*$ as in* (4.4)*, we have*

$$\text{cp-rank}(X) = 1 \quad \Leftrightarrow \quad \beta_1^2 = \alpha_1\alpha_2.$$

Proof. We have that cp-rank$(X) = 1$ if and only if there exists a $\mathbf{b} \in \mathbb{R}^3_+$ such that $X = \mathbf{b}\mathbf{b}^\mathsf{T}$. It is trivial that such a $\mathbf{b}$ is given by

$$\mathbf{b} = \frac{1}{\sqrt{\alpha_1}} \begin{pmatrix} \alpha_1 \\ \beta_1 \\ \beta_3 \end{pmatrix}.$$

From this we get that the cp-rank$(X) = 1$ if and only if

$$X = \begin{pmatrix} \alpha_1 & \beta_1 & \beta_3 \\ \beta_1 & \beta_1^2/\alpha_1 & \beta_1\beta_3/\alpha_1 \\ \beta_3 & \beta_1\beta_3/\alpha_1 & \beta_3^2/\alpha_1 \end{pmatrix}.$$

The forward implication is seen by comparing the required form of X to the original form of X.

For the reverse implication we note that if $\beta_1^2 = \alpha_1\alpha_2$, then from the requirements for complete positivity and the restrictions on X we have that

$$\alpha_2\alpha_3 \geq \beta_2^2 \geq \alpha_3\beta_1^2/\alpha_1 = \alpha_2\alpha_3,$$

implying that $\beta_2^2 = \alpha_2\alpha_3$. We also have that

$$\begin{aligned} 0 \leq \det(X) &= \alpha_1\alpha_2\alpha_3 + 2\beta_1\beta_2\beta_3 - \beta_1^2\alpha_3 - \beta_2^2\alpha_1 - \beta_3^2\alpha_2 \\ &= \alpha_1\alpha_2\alpha_3 + 2(\alpha_2\sqrt{\alpha_1\alpha_3})\beta_3 - \alpha_1\alpha_2\alpha_3 - \alpha_1\alpha_2\alpha_3 - \beta_3^2\alpha_2 \\ &= -\alpha_2(\beta_3 - \sqrt{\alpha_1\alpha_3})^2. \end{aligned}$$

This implies that X is in the required form. $\square$

Lemma 4.16. *For $X \in \mathcal{C}^*$ as in* (4.4)*, we have $\alpha_1\beta_2 - \beta_1\beta_3 \geq 0$.*

Proof. From Lemma 4.14 and the restrictions on X we have that $\beta_3^2 \leq \alpha_3\alpha_1$ and $0 \leq \alpha_3\beta_1^2 \leq \alpha_1\beta_2^2$. This implies that $\alpha_3\beta_1^2\beta_3^2 \leq \alpha_3\alpha_1^2\beta_2^2$, which gives the required result, due to X being strictly positive. $\square$

Lemma 4.17. *For $X \in \mathcal{C}^*$ as in* (4.4) *such that* $\text{cp-rank}(X) \neq 1$*, we have*

$$\text{cp-rank}(X) = 2 \quad \Leftrightarrow \quad \det(X) = 0.$$

Proof. From [BSM03, Theorem 3.2] we have that $\text{cp-rank}(X) = \operatorname{rank} X$. From the restrictions on X we have that $X \neq 0$, which, combined with the requirement that $\text{cp-rank}(X) \neq 1$, implies that $\operatorname{rank} X \geq 2$. From this we have

$$\begin{aligned} \text{cp-rank}(X) = 2 \quad &\Leftrightarrow \quad \operatorname{rank} X = 2 \\ &\Leftrightarrow \quad \operatorname{rank} X \neq 3 \\ &\Leftrightarrow \quad \det(X) = 0, \end{aligned}$$

which completes the proof. $\square$

From these lemmas we now present Algorithm 4.2 for testing whether a matrix X of the form given in (4.4) is completely positive and, if so, finding a minimal rank-one decomposition set for it. It can be seen that this method works in linear time with respect to the encoding length of the matrix X.

Algorithm 4.2 For testing whether a matrix $X \in \mathcal{S}^3$ of the form given in (4.4) is completely positive and, if so, finding a minimal rank-one decomposition for it.

Input: A matrix X of the form given in (4.4).

Output: Either "$X \notin \mathcal{C}^*$" or a set $\mathcal{B} \subset \mathbb{R}^3_+$ such that $|\mathcal{B}| = \text{cp-rank}(X)$ and $X = \sum_{\mathbf{b}\in\mathcal{B}} \mathbf{b}\mathbf{b}^\mathsf{T}$.

if $\beta_1^2 > \alpha_1\alpha_2$ **or** $\beta_2^2 > \alpha_2\alpha_3$ **or** $\beta_3^2 > \alpha_3\alpha_1$ **or** $\det(X) < 0$ **then**
  output "$X \notin \mathcal{C}^*$"
else
  initiate

$$\mathcal{B} = \left\{ \frac{1}{\sqrt{\alpha_1}} \begin{pmatrix} \alpha_1 \\ \beta_1 \\ \beta_3 \end{pmatrix} \right\}$$

  if $\beta_1^2 \neq \alpha_1\alpha_2$ **then**
    update

$$\mathcal{B} \leftarrow \mathcal{B} \cup \left\{ \sqrt{\frac{1}{\alpha_1(\alpha_1\alpha_2 - \beta_1^2)}} \begin{pmatrix} 0 \\ \alpha_1\alpha_2 - \beta_1^2 \\ \alpha_1\beta_2 - \beta_1\beta_3 \end{pmatrix} \right\}$$

    if $\det(X) \neq 0$ **then**
      update

$$\mathcal{B} \leftarrow \mathcal{B} \cup \left\{ \sqrt{\frac{\det(X)}{\alpha_1\alpha_2 - \beta_1^2}} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

    end if
  end if
  output $\mathcal{B}$
end if

4.5 Reducing Chain Lengths

Suppose that the matrix we wish to check for being completely positive gives the weighted-graph in Fig. 4.2. In this section we will see how we can reduce the length of the chain to give a smaller matrix while maintaining the property of whether the matrix is completely positive or not. For simplicity we shall view the matrices using their weighted-graph forms.

From considering the form of the rank-one decompositions when this graph is completely positive, we see that the graph gives a completely positive matrix if and only if there exists a y such that the chain in Fig. 4.1 and the weighted-graph in Fig. 4.3 give completely positive matrices.

Figure 4.2: We wish to analyse the matrix giving this weighted-graph to check for complete positivity, where the end points of the chain are distinct, the grey area represents an arbitrary structure in the graph, and $m > 3$.

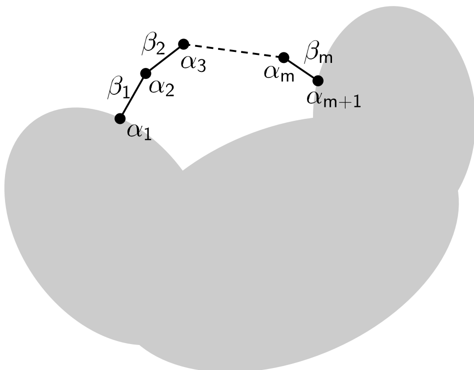
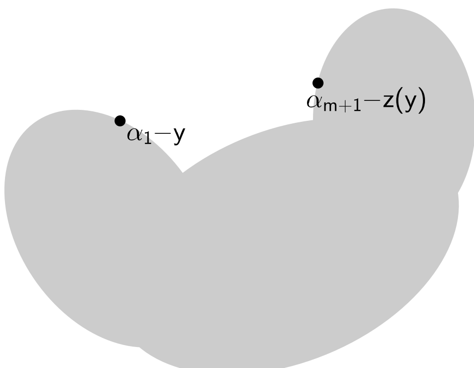


Figure 4.3: The weighted-graph in Fig. 4.2 gives a completely positive matrix if and only if there exists a y such that the chain in Fig. 4.1 and the weighted-graph below give completely positive matrices, where the grey area represents an arbitrary structure in the graph.

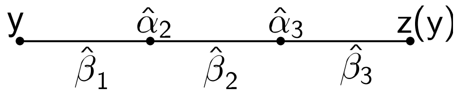


We now consider the chain in Fig. 4.1 in which the values of y and $z(y)$ are not fixed. We consider Method 4.10 on this chain. If the second step (checking $\lambda_i > 0$) finds that the chain cannot be part of a graph giving a completely positive matrix, then we are done. Otherwise we compare this chain to the chain in Fig. 4.4, where we set the values $\widehat{\alpha}_2$, $\widehat{\alpha}_3$, $\widehat{\beta}_1$, $\widehat{\beta}_2$, $\widehat{\beta}_3$, from the results of running the method for the original chain and where we pick arbitrary

$\gamma_1, \gamma_2 > 0$:

$$\begin{aligned} &\widehat{\alpha}_2 = \gamma_1 \lambda_{m-1}, \qquad \widehat{\alpha}_3 = \gamma_2 \mu_m, \qquad \widehat{\beta}_1 = \sqrt{\gamma_1 \mu_{m-1}}, \\ &\widehat{\beta}_2 = \sqrt{\gamma_1 \gamma_2 (\lambda_{m-1}\mu_m - \lambda_m \mu_{m-1})}, \qquad \widehat{\beta}_3 = \beta_m \sqrt{\gamma_2 \mu_{m-1}}. \end{aligned} \tag{4.5}$$

Figure 4.4: A chain we consider Method 4.10 working through with $n > 3$, $y > 0$, and the values for $\widehat{\alpha}_2$, $\widehat{\alpha}_3$, $\widehat{\beta}_1$, $\widehat{\beta}_2$ and $\widehat{\beta}_3$ given in (4.5).



We are free to pick whatever strictly positive values of γ_1 and γ_2 we wish without changing the theory. This freedom may, however, be able to be put to some advantages in reducing numerical difficulties in an algorithm, and it is recommended to pick values such that the order of magnitude on these vertices and edges is approximately that in the original weighted-graph.

From the results in Section 4.3 we can immediately see that all the vertices and edges in the chain have strictly positive values. We now consider Method 4.10 running through this chain.

i. We compute the values for $\widehat{\lambda}_i$, $\widehat{\mu}_i$, displayed in the following table:

i	$\widehat{\lambda}_i$	$\widehat{\mu}_i$
0	0	-1
1	1	0
2	$\gamma_1 \lambda_{m-1}$	$\gamma_1 \mu_{m-1}$
3	$\gamma_1 \gamma_2 \lambda_m \mu_{m-1}$	$\gamma_1 \gamma_2 \mu_m \mu_{m-1}$

ii. We can easily see that the values of $\widehat{\lambda}_i$ for $i \geq 2$ are strictly positive.

iii. We now have the following requirement on y and corresponding value for $z(y)$:

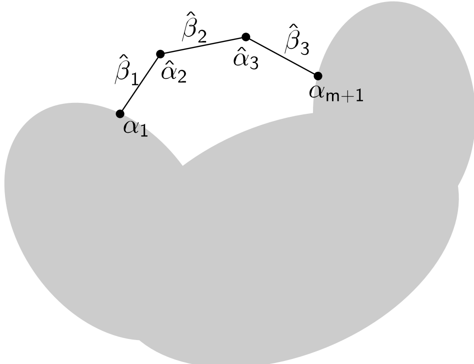
$$\begin{aligned} y &> \frac{\widehat{\mu}_3}{\widehat{\lambda}_3} = \frac{\mu_m}{\lambda_m}, \\ z(y) &= \frac{\widehat{\beta}_3^2 (\widehat{\lambda}_2 y - \widehat{\mu}_2)}{\widehat{\lambda}_3 y - \widehat{\mu}_3} = \frac{\beta_m^2 (\lambda_{m-1} y - \mu_{m-1})}{\lambda_m y - \mu_m}. \end{aligned}$$

Therefore, viewed from the end points, the chains in Figs. 4.1 and 4.4 are equivalent. Therefore, the graph in Fig. 4.2 is completely positive if and only if the following two conditions hold:

i. When computing the first two steps of Method 4.10 on the chain, we do not find that the chain cannot be part of a completely positive graph.

ii. The graph in Fig. 4.5 is completely positive, using the values given in (4.5).

Figure 4.5: Smaller graph for checking for complete positivity, which is equivalent to that in Fig. 4.2 using the results in Section 4.5.



We let X be the original matrix and Y be the matrix produced from our method. If we had a (minimal) rank-one decomposition set of Y then it would be a trivial task to convert this into a (minimal) rank-one decomposition set of X. We note that

$$\text{cp-rank}(X) = \text{cp-rank}(Y) + m - 3.$$

We finish this section by discussing the computation time of such a process. One simple method for applying this process is as follows, where we assume that no component submatrix is circular:

i. Find $\mathcal{W} = \{v \in \{1, \ldots, n\} \mid \text{The degree of } v \text{ in } G(X) \text{ is equal to } 2\}$.

ii. Find connected components of the subgraph of $G(X)$ induced by the vertices $\mathcal{W}$. Let these be denoted by the following with consecutive vertices being connected: $\{\{v_1^1, \ldots, v_{k_1}^1\}, \ldots, \{v_1^l, \ldots, v_{k_l}^l\}\}$.

iii. For all $i \in \{1, \ldots, l\}$ such that $k_i \geq 3$, do the following:

(a) Find $u, w \in \{1, \ldots, n\} \setminus \mathcal{W}$ such that $u, v_1^i, \ldots, v_{k_i}^i, w$ is a chain in $G(X)$.

(b) If $u \neq w$, then apply the method for reducing chain lengths to this chain.

(c) If $u = w$ and $k_i \geq 4$, then apply the method for reducing chain lengths to the chain $\{u, v_1^i, \ldots, v_{k_i}^i\}$.

It can be seen that this method would involve a linear number of calculations. In general we can not compute the square roots exactly, but if we are computing to a preset level of accuracy, then this method could be carried out in linear time.

4.6 Preprocessing

For a matrix $X \in \mathcal{S}^n$ we can now reduce the problem of checking whether it is completely positive and finding a (minimal) rank-one decomposition using the following linear-time method:

i. Check the matrix is nonnegative

ii. Check that whenever one of the matrix's on-diagonal entries is equal to zero, all of the off-diagonal entries in this row and column are also equal to zero.

iii. Reduce the problem to considering the maximal principal submatrix with strictly positive on-diagonal elements.

iv. Use Algorithm 4.1 to reduce the problem.

v. Split a matrix into its component submatrices (for example, with a breadth-first search).

vi. Use Section 4.1 to connect results from these submatrices to those for the original matrix.

vii. For each of these submatrices do the following:

  (a) If the resultant matrix is in $\mathcal{S}^3$, then use Algorithm 4.2 to process it.

  (b) Otherwise, if the resultant matrix is circular, use Method 4.13 to process it.

  (c) Otherwise use the method from Section 4.5 to reduce the chain lengths.

This method fully processes all component submatrices which have a maximum of one cycle. If all the component submatrices have a maximum of one cycle then this method determines whether the matrix is completely positive in linear-time and, if so, also outputs a minimal rank-one decomposition of it. Otherwise the method reduces the problem. As the method runs in linear-time and all known algorithms for computing the cp-rank in the general case run in exponential time [BR06], this is a very efficient preprocessor.

4.7 Number of minimal decompositions

Our method finds a single minimal rank-one decomposition set for a completely positive matrix such that every component submatrix has a maximum of one cycle. In this section we briefly look at how many minimal rank-one decomposition sets these matrices actually have. For simplicity we assume that the matrices are completely positive and connected and all the on-diagonal elements are strictly positive. We could then use Section 4.2 to extend these results to matrices where these assumptions do not hold.

If the cp-rank of a matrix $X \in \mathcal{S}^n$ is equal to one, then it is trivial to see that it has exactly one minimal rank-one decomposition set. Next we consider when the cp-rank of X is equal to two.

Theorem 4.18. *Let $X \in \mathcal{C}^{*n}$ be a connected matrix such that all the on-diagonal elements are strictly positive and* $\text{cp-rank}(X) = 2$. *Then the following hold:*

i. If there exists $i, j \in \{1, \dots, n\}$ such that $(X)_{ij} = 0$, then there is exactly one minimal rank-one decomposition set.

ii. If there does not exist $i, j \in \{1, \dots, n\}$ such that $(X)_{ij} = 0$, then there are infinitely many minimal rank-one decomposition sets.

Proof. This proof comes from considering the proof in [BSM03, Theorem 2.1]. A minimal rank-one decomposition set of X is of the form

$$\left\{ \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \right\}.$$

We now consider the ordered set of vectors

$$\mathcal{U} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\} \subset \mathbb{R}^2_+ \qquad \text{such that } \mathbf{u}_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix} \text{ for all } i.$$

X is then the gram matrix of these vectors; i.e. $(X)_{ij} = \langle \mathbf{u}_i, \mathbf{u}_j \rangle$ for all i, j. The minimal rank-one decomposition set is unique if and only if the ordered set $\mathcal{U}$ is unique up to a swapping of the coordinates of the $\mathbf{u}_i$'s. For $i, j = 1, \dots, n$ such that $i < j$, let θ_{ij} be the angle between the vectors $\mathbf{u}_i, \mathbf{u}_j \in \mathcal{U}$, such that $0 \le \theta_{ij} \le \pi$.

$$\frac{(X)_{ij}}{\sqrt{(X)_{ii}(X)_{jj}}} = \frac{\langle \mathbf{u}_i, \mathbf{u}_j \rangle}{\|\mathbf{u}_i\|_2 \|\mathbf{u}_j\|_2} = \cos\theta_{ij}.$$

We have that

$$X_{ij} > 0 \Leftrightarrow \theta_{ij} < \pi/2 \qquad \text{and} \qquad X_{ij} = 0 \Leftrightarrow \theta_{ij} = \pi/2.$$

Now let $\mathbf{u}_k, \mathbf{u}_l$ be the pair of vectors from $\mathcal{U}$ with maximal angle θ_{kl}. As $\mathcal{U} \subset \mathbb{R}^2_+$, we see that once the vectors $\mathbf{u}_k, \mathbf{u}_l$ are set, all the other vectors are uniquely defined and lie between them.

We now look at the two cases given in the theorem.

i. We have that $\theta_{kl} = \pi/2$, and so $\mathbf{u}_k$ and $\mathbf{u}_l$ lie on perpendicular axes. This implies that $\mathcal{U}$ is unique up to a swapping of coordinates and therefore there is exactly one minimal rank-one decomposition set.

ii. We have that $\theta_{kl} < \pi/2$. This gives us the freedom to rotate $\mathcal{U}$ while keeping it within $\mathbb{R}^2_+$, therefore there are infinitely many minimal rank-one decomposition sets. □

In Fig. 4.6 we now use this result to consider different types of matrices with the conditions given at the start of this section, i.e. completely positive and connected and all the diagonal elements strictly positive. The matrices we look at can, in fact, be easily extended to all the types of matrices that our method can check and decompose. For finding the cp-rank we simply consider how our method would work through this type of matrix.

Type	cp-rank	Number of minimal rank-one decomposition sets	Sketch of proof on number of minimal rank-one decomposition sets
Acyclic	$n-1$	1	From the form that a minimal rank-one decomposition set must take.
	n	∞	Consider stopping Algorithm 4.1 when exactly two (consecutive) indices are remaining. This effectively leaves us with a strictly positive matrix in $\mathcal{C}^{*2}$ to decompose with cp-rank equal to 2. Now use Theorem 4.18.
Circular, $n = 3$	1	1	As the matrix has cp-rank equal to 1.
	2	∞	Theorem 4.18.
	3	∞	Consider stopping Algorithm 4.2 after step 4, effectively leaving a strictly positive matrix in $\mathcal{C}^{*2}$ to decompose with cp-rank equal to 2. Now use Theorem 4.18.
Circular, $n > 3$	n	1 or 2	Method 4.13 results in either one or two solutions for y. Considering acyclic matrices with cp-rank equal to $n-1$, we see that each value of y gives exactly one minimal rank-one decomposition set. This was also found in [ZL00], where the authors found that the number of minimal rank-one decomposition sets of a completely positive circular matrix was dependent on the determinant of its comparison matrix.
$G(X)$ has exactly one cycle, which is of length equal to 3	$n-2$	1	From the form that a minimal rank-one decomposition set must take.
	$n-1$ or n	∞	Algorithm 4.1 followed by Algorithm 4.2 and considering circular matrices with $n = 3$ and cp-rank greater than or equal to 2.
$G(X)$ has exactly one cycle, which is of length greater than 3	n	1 or 2	Algorithm 4.1 followed by Method 4.13 gives the form that a minimal rank-one decomposition set must take, and then we consider circular matrices with $n > 3$.

Figure 4.6: Properties of minimal rank-one decompositions for some matrices which are completely positive and connected and have all on-diagonal entries strictly positive. Using the results of Section 4.2, we can extend these results to all the matrices which can be checked and decomposed by our method.

Part II

Geometry

Chapter 5

Proper Cones

In this part, we shall consider geometric properties of the copositive and completely positive cones. The first commonly known geometric property of these cones is that they are proper cones, where we recall that a proper cone is a cone which is closed, convex, pointed and full-dimensional. For the sake of completeness we shall give a proof of this here (without using properties of duality).

Lemma 5.1. *Both the copositive and the completely positive cones are full-dimensional.*

Proof. We note that $\mathcal{C}^{*n} \subseteq \mathcal{C}^n$, and thus we need only prove that the completely positive cone is full-dimensional.

If we consider the set of matrices

$$\{(\boldsymbol{e}_i + \boldsymbol{e}_j)(\boldsymbol{e}_i + \boldsymbol{e}_j)^\mathsf{T} \mid i \leq j, \quad i, j = 1, \ldots, n\} \subseteq \mathcal{C}^{*n}$$

then it can be observed that this is a set of $\frac{1}{2}n(n+1)$ linearly independent matrices, and thus the completely positive cone is full dimensional. $\square$

Lemma 5.2. *Both the copositive and completely positive cones are pointed.*

Proof. We note that $\mathcal{C}^{*n} \subseteq \mathcal{C}^n$, and thus we need only prove that the copositive cone is pointed.

We consider an arbitrary matrix $A \in \mathcal{C}^n \cap (-\mathcal{C}^n)$. For all $\mathbf{v} \in \mathbb{R}^n_+$ we have that $0 \leq \mathbf{v}^\mathsf{T} A \mathbf{v}$ and $0 \leq \mathbf{v}^\mathsf{T}(-A)\mathbf{v}$, implying that $\mathbf{v}^\mathsf{T} A \mathbf{v} = 0$. From this we immediately get that $A = 0$, completing the proof. $\square$

Lemma 5.3. *The copositive cone is a closed convex cone.*

Proof. The copositive cone is the intersection of infinitely many closed convex cones, and thus from Theorem 1.7 we see that it too is a closed convex cone. $\square$

Lemma 5.4. *The completely positive cone is a closed convex cone.*

Proof. From the definition it is trivial to observe that the completely positive cone is a convex cone.

We shall now prove that the completely positive cone is closed. We consider an arbitrary $A \in \operatorname{cl}(\mathcal{C}^{*n})$ and let $R = \langle A, I\rangle + 1$. It can then be observed that $A \in \operatorname{cl}\{X \in \mathcal{C}^{*n} \mid \langle X, I\rangle \leq R\}$ and

$$\begin{aligned}\{X \in \mathcal{C}^{*n} \mid \langle X, I\rangle \leq R\} &= \{\textstyle\sum_i \mathbf{a}_i\mathbf{a}_i^\mathsf{T} \mid \mathbf{a}_i \in \mathbb{R}^n_+ \text{ for all } i,\ \textstyle\sum_i \mathbf{a}_i^\mathsf{T}\mathbf{a}_i \leq R\}\\ &= \operatorname{conv}\{\mathbf{v}\mathbf{v}^\mathsf{T} \mid \mathbf{v} \in \mathbb{R}^n_+,\ \|\mathbf{v}\|_2^2 \leq R\}\\ &= \operatorname{conv}\widehat{\mathcal{L}}\end{aligned}$$

where

$$\widehat{\mathcal{L}} = \{\mathbf{v}\mathbf{v}^\mathsf{T} \mid \mathbf{v} \in \mathcal{L}\}, \qquad\qquad \mathcal{L} = \{\mathbf{v} \in \mathbb{R}^n_+ \mid \|\mathbf{v}\|_2^2 \leq R\}.$$

We have that $\mathcal{L}$ is a closed bounded set, which, from Lemma 1.15, implies that $\widehat{\mathcal{L}}$ is also a closed bounded set. This in turn implies that the convex hull of $\widehat{\mathcal{L}}$ is a closed bounded set. Therefore

$$A \in \operatorname{cl}\operatorname{conv}\widehat{\mathcal{L}} = \operatorname{conv}\widehat{\mathcal{L}} \subseteq \mathcal{C}^{*n},$$

which completes the proof. □

These lemmas can now be combined to give the following theorem.

Theorem 5.5. *The copositive and completely positive cones are both proper cones.*

Chapter 6

The Set of Zeros

In this chapter we consider a set referred to as the *Set of Zeros in the Nonnegative Orthant for a quadratic form*, or simply the *Set of Zeros* for short. This term, and the subsequent notation, was first introduced in the paper [Dic10], and builds on previous work from the paper [Dia62]. The definition of this set is as follows.

Definition 6.1. For $A \in \mathcal{S}^n$, we define the *Set of Zeros for* $\mathbf{x}^\mathsf{T} A\mathbf{x}$ *in the Nonnegative Orthant* by

$$\mathcal{V}^A := \{\mathbf{v} \in \mathbb{R}^n_+ \mid \mathbf{v}^\mathsf{T} A\mathbf{v} = 0\}.$$

Using this notation, we recall the following result from Theorem 1.1.

Theorem 6.2. *If $A \in \mathcal{C}$ and $\mathcal{V}^A \cap \mathbb{R}^n_{++} \neq \emptyset$, then $A \in \mathcal{S}_+$.*

This theorem can be generalised to give the following theorem:

Theorem 6.3. *Suppose the matrix $\widehat{A} \in \mathcal{C}^n$ and the vector $\widehat{\mathbf{x}} \in \mathbb{R}^n_+ \setminus \{\mathbf{0}\}$ can be partitioned as below, where $p \in \{1, \dots, n\}$, $A \in \mathcal{S}^p$, $B \in \mathbb{R}^{p\times(n-p)}$, $C \in \mathcal{S}^{n-p}$ and $\mathbf{x} \in \mathbb{R}^p_{++}$.*

$$\widehat{A} = \begin{pmatrix} A & B \\ B^\mathsf{T} & C \end{pmatrix}, \qquad \widehat{\mathbf{x}} = \begin{pmatrix} \mathbf{x} \\ \mathbf{0} \end{pmatrix}.$$

Then we have that $\widehat{\mathbf{x}} \in \mathcal{V}^{\widehat{A}}$ if and only if A is a positive semidefinite matrix and $\mathbf{x} \in \operatorname{Ker}(A)$, where $\operatorname{Ker}(A)$ denotes the kernel of A.

Proof. We have $\widehat{\mathbf{x}}^\mathsf{T} \widehat{A}\widehat{\mathbf{x}} = \mathbf{x}^\mathsf{T} A\mathbf{x}$, from which the reverse implication trivially follows. To prove the forward implication we first suppose that $\widehat{\mathbf{x}}^\mathsf{T} \widehat{A}\widehat{\mathbf{x}} = 0$, and thus $\mathbf{x}^\mathsf{T} A\mathbf{x} = 0$. From Theorem 1.1 we have that $A \in \mathcal{C}$, and, as $\mathbf{x}$ is strictly positive, Theorem 6.2 implies that $A \in \mathcal{S}_+$. Now, considering basic properties of positive semidefinite matrices, we get that $A\mathbf{x} = \mathbf{0}$. $\square$

We shall also consider the following related theorem, where this time we assume that A is positive semidefinite.

Theorem 6.4. *Suppose the matrix $\widehat{A} \in \mathcal{C}^n$ and the vector $\widehat{\mathbf{x}} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ can be partitioned as below, where $p \in \{1, \ldots, n\}$, $A \in \mathcal{S}^p_+$, $B \in \mathbb{R}^{p\times(n-p)}$, $C \in \mathcal{S}^{n-p}$ and $\mathbf{x} \in \mathbb{R}^p$.*

$$\widehat{A} = \begin{pmatrix} A & B \\ B^{\mathsf{T}} & C \end{pmatrix}, \qquad \widehat{\mathbf{x}} = \begin{pmatrix} \mathbf{x} \\ \mathbf{0} \end{pmatrix}.$$

Then we have that $\widehat{\mathbf{x}} \in \mathcal{V}^{\widehat{A}}$ if and only if $\mathbf{x} \in \mathbb{R}^p_+ \cap \operatorname{Ker}(A)$.

Proof. We have that $\widehat{\mathbf{x}}^{\mathsf{T}}\widehat{A}\widehat{\mathbf{x}} = 0$ if and only if $\mathbf{x}^{\mathsf{T}}A\mathbf{x} = 0$, which, as $A \in \mathcal{S}^p_+$, in turn holds if and only if $A\mathbf{x} = 0$. □

From these properties we immediately get the following two techniques connected to the set of zeros.

Method 6.5. Theorems 6.3 and 6.4 can be easily extended by considering permutations of the coordinate basis. From this we see that, for a given copositive matrix A, we can find $\mathcal{V}^A$ by first finding the maximal positive semidefinite principal submatrices of A and then considering their kernels. Applying this method for a given copositive matrix A, we find that its set of zeros in the nonnegative orthant is of the form

$$\mathcal{V}^A = \bigcup_{i=1}^{m} \operatorname{conic} \mathcal{X}_i,$$

where $\mathcal{X}_i \subset \mathbb{R}^n_+$ is a finite set for all i. Each $\mathcal{X}_i$ relates to a set of generators for the exposed rays of the intersection of the nonnegative orthant with the kernel of a maximal positive semidefinite principal submatrix.

Method 6.6. We can also partially reverse the process in the previous method. Given a finite set $\mathcal{V} \subset \mathbb{R}^n_+$, we can find necessary conditions on a matrix $A \in \mathcal{C}$ in order to have $\mathcal{V} \subset \mathcal{V}^A$. These necessary conditions are in terms of certain principal submatrices being positive semidefinite and containing certain vectors in their kernels.

We finish this chapter by considering the following useful lemma in connection to the set of zeros.

Lemma 6.7. *For $U = \sum_{i=1}^m \mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}$ such that $\mathbf{u}_i \in \mathbb{R}^n_+$ for all $i = 1, \ldots, m$, and $A \in \mathcal{C}$, we have that $\langle U, A\rangle = 0$ if and only if $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\} \subseteq \mathcal{V}^A$.*

Proof. We have that $\langle U, A\rangle = \left\langle \sum_{i=1}^m \mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}, A\right\rangle = \sum_{i=1}^m \mathbf{u}_i^{\mathsf{T}}A\mathbf{u}_i$.
From the definition of a copositive matrix, $0 \leq \mathbf{u}_i^{\mathsf{T}}A\mathbf{u}_i$ for all i.
Therefore

$$0 = \langle U, A\rangle \quad\Leftrightarrow\quad 0 = \mathbf{u}_i^{\mathsf{T}}A\mathbf{u}_i \text{ for all } i \quad\Leftrightarrow\quad \{\mathbf{u}_1, \ldots, \mathbf{u}_m\} \subseteq \mathcal{V}^A. \qquad \square$$

Chapter 7

Interior of Copositive and Completely Positive Cones*

7.1 Introduction

In this chapter we will consider the interiors of the copositive and completely positive cones. These results primarily come from the paper [Dic10], which was an extension of the paper [DS08], and has itself been extended to set-semidefinite cones in the paper [GS11].

When analysing the interiors of proper cones, the following theorem is highly useful.

Theorem 7.1 ([Ber73])**.** *Let $\mathcal{K} \subseteq \mathcal{S}^n$ be a proper cone. Then we have that*

$$\operatorname{int}(\mathcal{K}^*) = \{X \in \mathcal{S}^n \mid \langle X, Y\rangle > 0 \textit{ for all } Y \in \mathcal{K} \setminus \{0\}\}.$$

From this we immediately get the following well-known characterisation for the interior of the copositive cone:

Theorem 7.2. *We have that*

$$\operatorname{int}(\mathcal{C}^n) = \{X \in \mathcal{S}^n \mid \mathbf{v}^\mathsf{T} X\mathbf{v} > 0 \textit{ for all } \mathbf{v} \in \mathbb{R}^n_+ \setminus \{0\}\}.$$

The interior of the completely positive cone is however a lot more complicated. This was first analysed by Dür and Still [DS08], who found that

$$\operatorname{int}(\mathcal{C}^*) = \left\{ AA^\mathsf{T} \;\middle|\; A = [A_1|A_2] \text{ with } A_1 > 0 \text{ nonsingular, } A_2 \geq 0 \right\}.$$

---

*Submitted as:

[Dic10] P.J.C. Dickinson. An improved characterisation of the interior of the completely positive cone. *Electronic Journal of Linear Algebra*, 20:723–729, 2010.

Alternatively, this can be denoted as

$$\operatorname{int}(\mathcal{C}^*) = \left\{ \sum_{i=1}^{m} \mathbf{a}_i \mathbf{a}_i^{\mathsf{T}} \;\middle|\; \begin{array}{l} m \geq n, \quad \mathbf{a}_i \in \mathbb{R}^n_+ \text{ for all } i, \\ \{\mathbf{a}_1, \ldots, \mathbf{a}_n\} \subseteq \mathbb{R}^n_{++}, \\ \operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = \mathbb{R}^n \end{array} \right\}.$$

This characterisation of the interior is useful in that we can construct any matrix in the interior from it, and any matrix in the interior can be decomposed into this form. However, from Theorem 4.1, we see that any completely positive matrix in the interior can be decomposed in a way that is not of this form (provided $n > 1$). An alternative explicit example is below.

Example 7.3. We consider the matrix $X_\alpha = (I + \alpha E) \in \mathcal{S}^n$ with $n > 1$ and $\alpha > 0$.

Letting $\beta = \frac{1}{n}\left(\sqrt{1 + \alpha n} - 1\right) > 0$, we have that $(I + \beta E)$ is a nonsingular matrix, with all entries strictly positive, such that $X_\alpha = (I + \beta E)(I + \beta E)^{\mathsf{T}}$, and thus $X_\alpha \in \operatorname{int} \mathcal{C}^*$.

However, we also have that $X_\alpha = \alpha \boldsymbol{e}\boldsymbol{e}^{\mathsf{T}} + \sum_{i=1}^n \boldsymbol{e}_i \boldsymbol{e}_i^{\mathsf{T}}$, and from this decomposition we would not be able to see that X_α is in the interior of the completely positive cone using the characterisation above.

How to tell if an arbitrary completely positive matrix is in the interior or not from a general rank-one decomposition of it is still an open question. However, in this chapter we shall prove the following theorem, which provides alternative characterisations of the interior.

Theorem 7.4. *For $n \in \mathbb{Z}_{++}$ we consider the sets:*

$$\begin{aligned} \mathcal{M}_1 &= \left\{ \sum_{i=1}^{m} \mathbf{a}_i \mathbf{a}_i^{\mathsf{T}} \;\middle|\; \begin{array}{l} \mathbf{a}_i \in \mathbb{R}^n_{++} \textit{ for all } i, \\ \operatorname{span}\{\mathbf{a}_1, \ldots \mathbf{a}_m\} = \mathbb{R}^n \end{array} \right\}, \\ &= \{AA^{\mathsf{T}} \mid A > 0,\ \operatorname{rank} A = n\}, \end{aligned}$$

$$\begin{aligned} \mathcal{M}_2 &= \left\{ \sum_{i=1}^{m} \mathbf{a}_i \mathbf{a}_i^{\mathsf{T}} \;\middle|\; \begin{array}{l} m \geq n, \quad \mathbf{a}_i \in \mathbb{R}^n_+ \textit{ for all } i, \\ \{\mathbf{a}_1, \ldots, \mathbf{a}_n\} \subseteq \mathbb{R}^n_{++}, \\ \operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = \mathbb{R}^n \end{array} \right\} \\ &= \left\{ AA^{\mathsf{T}} \;\middle|\; A = [A_1 | A_2] \textit{ with } A_1 > 0 \textit{ nonsingular, } A_2 \geq 0 \right\}, \end{aligned}$$

$$\begin{aligned} \mathcal{M}_3 &= \left\{ \sum_{i=1}^{m} \mathbf{a}_i \mathbf{a}_i^{\mathsf{T}} \;\middle|\; \begin{array}{l} \mathbf{a}_1 \in \mathbb{R}^n_{++},\ \mathbf{a}_i \in \mathbb{R}^n_+ \textit{ for all } i, \\ \operatorname{span}\{\mathbf{a}_1, \ldots \mathbf{a}_m\} = \mathbb{R}^n \end{array} \right\}, \\ &= \left\{ AA^{\mathsf{T}} \mid \operatorname{rank} A = n,\ A = [\mathbf{a}|B],\ \mathbf{a} \in \mathbb{R}^n_{++},\ B \geq 0 \right\}. \end{aligned}$$

We then have that $\operatorname{int} \mathcal{C}^{*n} = \mathcal{M}_1 = \mathcal{M}_2 = \mathcal{M}_3$.

This was originally proven in the paper [Dic10]. Since then the author has found a new proof of this. In this chapter both proofs are included as they provide different insights into the problem.

7.2 Original Proof

Considering the sets $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ from Theorem 7.4, it is trivial to see that $\mathcal{M}_1 \subseteq \mathcal{M}_2 \subseteq \mathcal{M}_3$. In this section we shall show that $\mathcal{M}_3 \subseteq \operatorname{int} \mathcal{C}^* \subseteq \mathcal{M}_1$, which will prove Theorem 7.4.

We begin by noting that, as the copositive and completely positive cones are duals of each other, Theorem 7.1 gives us the following result.

Lemma 7.5. *For an arbitrary $U \in \mathcal{C}^*$, we have that $U \in \operatorname{bd}(\mathcal{C}^*)$ if and only if there exists $X \in \mathcal{C} \setminus \{0\}$ such that $\langle U, X \rangle = 0$.*

We can now combine this with one of the properties of the Set of Zeros in the Nonegative Orthant (Lemma 6.7) to get the following.

Lemma 7.6. *Consider $U = \sum_{i=1}^m \mathbf{a}_i \mathbf{a}_i^\mathsf{T}$ such that $\mathbf{a}_i \in \mathbb{R}^n_+$ for all i. Then we have that $U \in \operatorname{bd}(\mathcal{C}^*)$ if and only if there exists $X \in \mathcal{C} \setminus \{0\}$ such that $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\} \subseteq \mathcal{V}^X$.*

We will now use this to prove one of the inclusion relations.

Theorem 7.7. *We have $\mathcal{M}_3 \subseteq \operatorname{int} \mathcal{C}^*$.*

Proof. We consider a matrix $U \in \mathcal{M}_3$. There exists $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n_+$ such that $\mathbf{a}_1 \in \mathbb{R}^n_{++}$ and $\operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\} = \mathbb{R}^n$ and $U = \sum_{i=1}^m \mathbf{a}_i \mathbf{a}_i^\mathsf{T}$. We have that $U \in \mathcal{C}^*$, and suppose for the sake of contradiction that $U \in \operatorname{bd}(\mathcal{C}^*)$. From Lemma 7.6, this implies that there exists $X \in \mathcal{C} \setminus \{0\}$ such that we have $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\} \subseteq \mathcal{V}^X$. Therefore $\mathbf{a}_1 \in \mathcal{V}^X \cap \mathbb{R}^n_{++}$, and so, from Theorem 6.2, we get $X \in \mathcal{S}_+$. We now have that

$$X \in \mathcal{S}_+, \ \mathbf{a}_i^\mathsf{T} X \mathbf{a}_i = 0 \Rightarrow X \mathbf{a}_i = 0$$

Therefore $X\mathbf{a}_i = 0$ for all i. This then implies that we have $X\mathbf{v} = 0$ for all $\mathbf{v} \in \operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\} = \mathbb{R}^n$, and thus $X = 0$, which is a contradiction. $\square$

In order to prove the other inclusion relation, we shall first show that the spanning constraint is in fact a necessary condition for the rank-one decomposition of a matrix in the interior of the completely positive cone.

Lemma 7.8. *Consider $U = \sum_{i=1}^m \mathbf{a}_i \mathbf{a}_i^\mathsf{T}$ such that $\mathbf{a}_i \in \mathbb{R}^n_+$ for all i and $\operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\} \neq \mathbb{R}^n$. Then $U \in \operatorname{bd}(\mathcal{C}^*)$.*

Proof. Recall that $\mathcal{C}^* \subseteq \mathcal{S}_+$, and note that $\operatorname{span}\{\mathbf{a}_1, \dots, \mathbf{a}_m\} \neq \mathbb{R}^n$ implies that $\operatorname{rank} U < n$. From this we get that $U \in \operatorname{bd}(\mathcal{S}_+)$, which in turn implies that $U \in \operatorname{bd}(\mathcal{C}^*)$. $\square$

We now consider the Krein-Milman Theorem, as stated in [BSM03, page 45].

Theorem 7.9 (Krein-Milman Theorem)**.** *If T is a set of extreme vectors of a closed convex cone K which generate all the extreme rays of K, then*

$$K = \operatorname{cl} \operatorname{conic} T$$

From this we get the following lemma.

Lemma 7.10. *If $\mathcal{D}$ is a convex cone contained in a proper cone $\mathcal{K}$ such that all the extreme rays of $\mathcal{K}$ are contained in* $\operatorname{cl}(\mathcal{D})$*, then* $\operatorname{int}(\mathcal{K}) \subseteq \mathcal{D}$*.*

Proof. It can be seen from the Krein-Milman Theorem that $\mathcal{K} = \operatorname{cl}(\mathcal{D})$.

Now suppose for the sake of contradiction that there exists $\mathbf{x} \in \operatorname{int}(\mathcal{K}) \setminus \mathcal{D}$. It is a standard result that as $\mathcal{D}$ is convex there exists a hyperplane through $\mathbf{x}$ giving a closed halfspace H such that $\mathcal{D} \subseteq H$. This implies that $\operatorname{cl}(\mathcal{D}) \subseteq H$. However we also have that $\mathbf{x} \in \mathcal{K} = \operatorname{cl}(\mathcal{D}) \subseteq H$. The fact that the hyperplane goes through $\mathbf{x}$ implies that $\mathbf{x} \in \operatorname{bd}(H)$, and thus $\mathbf{x} \in \operatorname{bd}(\mathcal{K})$, which is a contradiction. $\square$

This then gives us the other inclusion relation.

Theorem 7.11. *We have* $\operatorname{int} \mathcal{C}^* \subseteq \mathcal{M}_1$*.*

Proof. Let $\mathcal{D} = \{AA^\mathsf{T} \mid A > 0\} \cup \{0\}$. It is not difficult to see that this is a convex cone which is contained in the completely positive cone.
From [BSM03, page 71] we have that the completely positive cone is a proper cone, with the extreme rays being the matrices $\mathbf{b}\mathbf{b}^\mathsf{T}$ where $\mathbf{b} \in \mathbb{R}^n_+ \setminus \{\mathbf{0}\}$. These are obviously members of $\operatorname{cl}(\mathcal{D})$. Therefore, from Lemma 7.10, we have that $\operatorname{int}(\mathcal{C}^*) \subseteq \mathcal{D}$.
Finally, for an arbitrary $A > 0$, using Theorem 7.7 and Lemma 7.8, we have that

$$AA^\mathsf{T} \in \operatorname{int}(\mathcal{C}^*) \Leftrightarrow \operatorname{rank} A = n.$$

$\square$

7.3 New Proof

As previously stated, considering the sets $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ from Theorem 7.4, it is trivial to see that $\mathcal{M}_1 \subseteq \mathcal{M}_2 \subseteq \mathcal{M}_3$. In this section we shall give a direct proof that $\mathcal{M}_3 \subseteq \mathcal{M}_1$. Then using the result that $\operatorname{int} \mathcal{C}^* = \mathcal{M}_2$ from [DS08], this will prove Theorem 7.4.

Using the following two technical lemmas, it is trivial to see that we can take a rank-one decomposition set in the form required for $\mathcal{M}_3$ and transform it into the form required for $\mathcal{M}_1$. This then implies that the required inclusion relation holds.

Lemma 7.12. *Let $U = \sum_{i=1}^{m} \mathbf{a}_i \mathbf{a}_i^\mathsf{T}$ such that $\mathbf{a}_i \in \mathbb{R}^n_+ \setminus \{0\}$ for all i. Then we have that*

$$\operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\} = \mathbb{R}^n \qquad \Leftrightarrow \qquad \operatorname{rank} U = n.$$

Proof. We let $A \in \mathbb{R}^{n \times m}$ such that the columns of A are given by the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$. We have $U = AA^\mathsf{T}$, and that $\operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\} = \mathbb{R}^n$ if and only if $\operatorname{rank} A = n$. We now recall the well-known result that $\operatorname{rank} AA^\mathsf{T} = \operatorname{rank} A$ (see for example [Ber09, Corollary 2.5.1]), which completes the proof. $\square$

Lemma 7.13. *Consider two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n_+$ such that $\mathbf{a} \in \mathbb{R}^n_{++}$. Then there exist vectors $\mathbf{c}, \mathbf{d} \in \mathbb{R}^n_{++}$ such that $\mathbf{a}\mathbf{a}^\mathsf{T} + \mathbf{b}\mathbf{b}^\mathsf{T} = \mathbf{c}\mathbf{c}^\mathsf{T} + \mathbf{d}\mathbf{d}^\mathsf{T}$.*

Proof. For $\theta \in \mathbb{R}$, we define

$$\begin{aligned} \mathbf{a}_\theta &:= \mathbf{a} \cos\theta - \mathbf{b} \sin\theta, \\ \mathbf{b}_\theta &:= \mathbf{b} \cos\theta + \mathbf{a} \sin\theta. \end{aligned}$$

It can be observed that

$$\mathbf{a}\mathbf{a}^\mathsf{T} + \mathbf{b}\mathbf{b}^\mathsf{T} = \mathbf{a}_\theta \mathbf{a}_\theta^\mathsf{T} + \mathbf{b}_\theta \mathbf{b}_\theta^\mathsf{T}.$$

Now note that as $\mathbf{a} \in \mathbb{R}^n_{++}$ and $\mathbf{b} \in \mathbb{R}^n_+$, by picking θ to be positive and sufficiently small, we can ensure that $\mathbf{a}_\theta, \mathbf{b}_\theta \in \mathbb{R}^n_{++}$. We now let $\mathbf{c}$ and $\mathbf{d}$ equal these vectors to complete the proof. $\square$

Remark 7.14. When transforming a rank-one decomposition set from one in the form required for $\mathcal{M}_3$ to one in the form required for $\mathcal{M}_1$, the cardinality of the set is not altered.

7.4 Concluding Remarks

In this chapter we have shown some alternative characterisations for the interior of the completely positive cone. This includes both a fairly relaxed characterisation ($\mathcal{M}_3$) and a very restrictive one ($\mathcal{M}_1$).

It is trivial to see that for $n \leq 2$, we have that $\mathcal{M}_3$ characterises all rank-one decompositions of matrices in $\operatorname{int} \mathcal{C}^{*n}$. In the following lemma we shall return to example 7.3 and show that for $n > 2$, some matrices in $\operatorname{int} \mathcal{C}^{*n}$ have all their rank-one decompositions given in $\mathcal{M}_3$, but some have rank-one decompositions which are not given in $\mathcal{M}_3$.

Lemma 7.15. *We consider the matrix* $X_\alpha = (I + \alpha E) \in \mathcal{S}^n$ *with* $n > 2$ *and* $\alpha > 0$*, and recall that in example 7.3 it was shown that* $X_\alpha \in \operatorname{int} \mathcal{C}^*$*. We have that there exists a set of vectors* $\{\mathbf{a}_1, \dots, \mathbf{a}_m\} \subseteq \mathbb{R}^n_+ \setminus \mathbb{R}^n_{++}$ *such that* $X_\alpha = \sum_{i=1}^m \mathbf{a}_i \mathbf{a}_i^\mathsf{T}$ *if and only if* $\alpha \leq n - 2$.

Proof. The reverse implication comes trivially from noting that

$$X_\alpha = \frac{\alpha}{n-2} \sum_{i=1}^n (\boldsymbol{e} - \boldsymbol{e}_i)(\boldsymbol{e} - \boldsymbol{e}_i)^\mathsf{T} + \left(1 - \frac{\alpha}{n-2}\right) \sum_{i=1}^n \boldsymbol{e}_i \boldsymbol{e}_i^\mathsf{T}.$$

We shall now prove the forward implication. We consider $\alpha > 0$ such that there exists a set $\{\mathbf{a}_1, \dots, \mathbf{a}_m\} \subseteq \mathbb{R}^n_+ \setminus \mathbb{R}^n_{++}$ with $X_\alpha = \sum_{i=1}^m \mathbf{a}_i \mathbf{a}_i^\mathsf{T}$.

From Lemma 1.35 it can be seen that the matrix $\big((n-1)I - E\big) \in \mathcal{S}^{n-1}$ is copositive. Therefore, if we now consider the matrix $Z = \big((n-1)I - E\big) \in \mathcal{S}^n$, then all principal submatrices of this with order less than or equal to $n - 1$ are copositive. This implies that $0 \leq \mathbf{a}_i^\mathsf{T} Z \mathbf{a}_i$ for all i, which then gives the following, completing the proof:

$$0 \leq \sum_{i=1}^m \mathbf{a}_i^\mathsf{T} Z \mathbf{a}_i = \langle I + \alpha E, Z \rangle = n\big((n-2) - \alpha\big). \qquad \square$$

Although the relaxed characterisation does not solve the problem of using a general rank-one decomposition to tell if a matrix is in the interior of the completely positive cone, it is nonetheless a fairly relaxed characterisation of the interior.

Chapter 8

Facial Structure*

8.1 Geometry of General Proper Cones

In this chapter we will consider the facial structure of the copositive and completely positive cones. In order to do this we will start by considering properties for general proper cones, where we recall that a proper cone is a cone which is closed, convex, pointed and full-dimensional. Although we shall consider the properties for proper cones contained in the space of real vectors, these definitions and results can trivially be extended to proper cones in spaces which are isomorphic to the real space, for example the set of symmetric matrices, which is the space that the copositive and completely cones sit in.

We begin with some definitions, with the definitions for a face, an exposed face, an exposed ray and an extreme ray being equivalent to those used in [Roc70, Section 18].

Definition 8.1. A face of a closed convex set $\mathcal{L} \subseteq \mathbb{R}^n$ is a convex subset $\mathcal{F} \subseteq \mathcal{L}$ such that every closed line segment in $\mathcal{L}$ with a relative interior point in $\mathcal{F}$ has both end points in $\mathcal{F}$. A facet of a closed convex set $\mathcal{L}$ is a face of the set with dimension equal to $\dim \mathcal{L} - 1$. An extreme point of a closed convex set $\mathcal{L}$ is a face of the set with dimension equal to zero.

Definition 8.2. Let $\mathcal{L}$ be a closed convex set in $\mathbb{R}^n$ and $\emptyset \neq \mathcal{F} \subset \mathcal{L}$. $\mathcal{F}$ is an exposed face of $\mathcal{L}$ if it is the intersection of $\mathcal{L}$ and a non-trivial supporting hyperplane, i.e. if there exists $\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, $\beta \in \mathbb{R}$ such that we have $\mathcal{L} \subseteq \{\mathbf{x} \in \mathbb{R}^n \mid \langle \mathbf{x}, \mathbf{a} \rangle \geq \beta\}$ and $\mathcal{F} = \{\mathbf{x} \in \mathcal{L} \mid \langle \mathbf{x}, \mathbf{a} \rangle = \beta\}$. An exposed point of a closed convex set $\mathcal{L}$ is an exposed face of the set with dimension equal to zero. (Rockafellar also refers to $\mathcal{L}$ and $\emptyset$ as exposed faces, however we shall exclude

*Submitted as:

[Dic11] P.J.C. Dickinson. Geometry of the copositive and completely positive cones. *Journal of Mathematical Analysis and Applications*, 380(1):377–395, 2011.

these. In much of the literature, for example [Brø83], these faces are called improper exposed faces whilst the exposed faces that we will be considering are called (proper) exposed faces.)

Remark 8.3. Every exposed face is also a face.

Theorem 8.4. *Every face of a full dimensional closed convex set $\mathcal{L}$ which is not equal to $\mathcal{L}$ is contained within an exposed face.*

Proof. If $\mathcal{F}_1 \neq \mathcal{L}$ is an arbitrary face of $\mathcal{L}$, then we have $\mathcal{F}_1 \subseteq \mathrm{bd}(\mathcal{L})$. Let $\mathbf{x}$ be in the relative interior of $\mathcal{F}_1$. By the supporting hyperplane theorem there exists an exposed face $\mathcal{F}_2$ such that $\mathbf{x} \in \mathcal{F}_2$. Therefore $\mathcal{F}_2 \cap \mathrm{reint}(\mathcal{F}_1) \neq \emptyset$, and so from [Roc70, Theorem 18.1] we have $\mathcal{F}_1 \subseteq \mathcal{F}_2$. □

Definition 8.5. A face $\mathcal{F}_1$ is a maximal face of a full dimensional closed convex set $\mathcal{L}$ if $\mathcal{F}_1 \neq \mathcal{L}$ and there does not exist a face $\mathcal{F}_2 \neq \mathcal{L}$ such that $\mathcal{F}_1 \subset \mathcal{F}_2$.

Remark 8.6. From Theorem 8.4 it can be immediately seen that every maximal face is also an exposed face.

Remark 8.7. The maximal faces of a polyhedron are its facets and the maximal faces of an n-sphere are the points on its boundary.

The following theorem contributes towards our motivation for looking at the set of maximal faces as it means that the hyperplanes giving maximal faces are desirable in a cutting plane algorithm.

Theorem 8.8. *Let $\mathbb{M}$ be the set of maximal faces of a full-dimensional closed convex set $\mathcal{L}$, and let $\mathbb{S}$ be an arbitrary set of its faces, none of which are equal to the complete set. Then,*

$$\bigcup_{\mathcal{F}\in\mathbb{S}} \mathcal{F} = \mathrm{bd}(\mathcal{L}) \quad \Leftrightarrow \quad \mathbb{M} \subseteq \mathbb{S}.$$

Proof. ($\Leftarrow$) By the supporting hyperplane theorem, every point on $\mathrm{bd}(\mathcal{L})$ is a member of an exposed face and therefore is also a member of a maximal face. This implies that $\mathrm{bd}(\mathcal{L}) = \bigcup_{\mathcal{F}\in\mathbb{M}} \mathcal{F}$.
($\Rightarrow$) Suppose for the sake of contradiction that $\bigcup_{\mathcal{F}\in\mathbb{S}} \mathcal{F} = \mathrm{bd}(\mathcal{L})$ and there exists a maximal face $\mathcal{F}_1 \notin \mathbb{S}$. We consider an arbitrary point $\mathbf{x} \in \mathrm{reint}(\mathcal{F}_1)$. We have that $\mathbf{x} \in \mathrm{bd}(\mathcal{L})$, therefore there exists a face $\mathcal{F}_2 \in \mathbb{S}$ such that $\mathbf{x} \in \mathcal{F}_2$. This implies that $\mathcal{F}_2 \cap \mathrm{reint}(\mathcal{F}_1) \neq \emptyset$ and so from [Roc70, Theorem 18.1] we get that $\mathcal{F}_1 \subset \mathcal{F}_2$, implying that $\mathcal{F}_1$ can not be a maximal face. □

We now switch our focus to rays, in particular the exposed and extreme rays. If we consider an arbitrary $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, then the ray generated by $\mathbf{x}$ is defined to be the set $\{\alpha\mathbf{x} \mid \alpha \geq 0\}$. We say that $\mathbf{x}$ is a generator of this ray.

Definition 8.9. $\mathbf{x} \in \mathcal{K} \setminus \{\mathbf{0}\}$ generates an exposed ray of a proper cone $\mathcal{K}$ if there exists an exposed face $\mathcal{F}$ of $\mathcal{K}$ such that

$$\mathcal{F} = \{\alpha\mathbf{x} \mid \alpha \geq 0\}.$$

We write $\mathrm{Exp}(\mathcal{K})$ for the set of elements generating exposed rays of $\mathcal{K}$.

Definition 8.10. $\mathbf{x} \in \mathcal{K} \setminus \{\mathbf{0}\}$ generates an extreme ray of a proper cone $\mathcal{K}$ if

$$\mathbf{y}, \mathbf{z} \in \mathcal{K},\ \mathbf{y} + \mathbf{z} = \mathbf{x} \ \Rightarrow\ \mathbf{y}, \mathbf{z} \in \{\alpha\mathbf{x} \mid \alpha \geq 0\}.$$

We write $\mathrm{Ext}(\mathcal{K})$ for the set of elements generating extreme rays of $\mathcal{K}$.

Theorem 8.11 (Straszewicz's theorem, see [Roc70, Theorem 18.6])**.** *For a closed convex set, the set of exposed points is a dense subset of the set of extreme points.*

This can be extended to rays of a proper cone, giving the following.

Theorem 8.12. *For a proper cone $\mathcal{K}$,*

$$\mathrm{Exp}(\mathcal{K}) \subseteq \mathrm{Ext}(\mathcal{K}) \subseteq \mathrm{cl}(\mathrm{Exp}(\mathcal{K})).$$

Proof. As $\mathcal{K}$ is a closed pointed cone there exists a bounded base of it, given by $\mathcal{B} = \mathcal{H} \cap \mathcal{K}$, for some hyperplane $\mathcal{H}$. It can be seen that $\mathbf{x} \in \mathrm{Ext}(\mathcal{K})$ ($\mathbf{x} \in \mathrm{Exp}(\mathcal{K})$) if and only if there exists $\alpha > 0$ such that $\alpha\mathbf{x}$ is an extreme (exposed) point of $\mathcal{B}$. We now consider Straszewicz's theorem to get the desired result. $\square$

For the rest of this section we will consider how the faces of a proper cone are related to points in its dual.

Theorem 8.13. *$\mathcal{F}$ is an exposed face of a proper cone $\mathcal{K}$ if and only if there exists an $\mathbf{a} \in \mathcal{K}^* \setminus \{\mathbf{0}\}$ such that*

$$\mathcal{F} = \mathcal{F}(\mathcal{K}, \mathbf{a}) := \{\mathbf{x} \in \mathcal{K} \mid \langle \mathbf{x}, \mathbf{a} \rangle = 0\}.$$

Proof. From [BV04, page 51] we have that

> $\mathbf{y} \in \mathcal{K}^*$ if and only if $-\mathbf{y}$ is the normal of a hyperplane that supports $\mathcal{K}$ at the origin.

If we now consider any nonzero point in a face of $\mathcal{K}$, then from the definition of a face we get that the ray generated by this point is also contained within the face. This implies that all nonempty faces of a proper cone contain the origin. Combining these two facts gives us the required result. $\square$

Using the following observation we now get a similar result relating the maximal faces of a proper cone to the extreme rays in its dual. This lemma can be immediately seen from the definition of $\mathcal{F}(\mathcal{K}, \mathbf{a})$ in Theorem 8.13 and the definition of the dual, so it is presented without proof.

Lemma 8.14. *For $\{\mathbf{a}_1, \dots, \mathbf{a}_m\} \subset \mathcal{K}^*$, we have that*

$$\mathcal{F}\left(\mathcal{K}, \textstyle\sum_{i=1}^m \mathbf{a}_i\right) = \bigcap_{i=1}^m \mathcal{F}(\mathcal{K}, \mathbf{a}_i),$$

where we extend the definition of $\mathcal{F}(\mathcal{K}, \mathbf{a})$ such that $\mathcal{F}(\mathcal{K}, \mathbf{0}) := \mathcal{K}$.

Theorem 8.15. *If $\mathcal{F}$ is a maximal face of a proper cone $\mathcal{K}$ then there exists an $\mathbf{a} \in \mathrm{Ext}(\mathcal{K}^*)$ such that $\mathcal{F} = \mathcal{F}(\mathcal{K}, \mathbf{a})$.*

Proof. Let $\mathcal{F}$ be a maximal face of $\mathcal{K}$. Then $\mathcal{F}$ is an exposed face, and so by Theorem 8.13, $\mathcal{F} = \mathcal{F}(\mathcal{K}, \mathbf{a})$ for some $\mathbf{a} \in \mathcal{K}^* \setminus \{\mathbf{0}\}$. It is a well-known result that $\mathbf{a}$ can be decomposed as $\mathbf{a} = \sum_{j \in \mathcal{J}} \mathbf{a}_j$, where $\{\mathbf{a}_j\}_{j \in \mathcal{J}} \subseteq \mathrm{Ext}(\mathcal{K}^*)$. This is in fact an extension of the Krein-Milman theorem [KM40]. Therefore

$$\begin{aligned} \mathcal{F} &= \mathcal{F}\left(\mathcal{K}, \textstyle\sum_{j \in \mathcal{J}} \mathbf{a}_j\right) \\ &= \bigcap_{j \in \mathcal{J}} \mathcal{F}(\mathcal{K}, \mathbf{a}_j) \qquad \text{(Lemma 8.14)} \\ &\subseteq \mathcal{F}(\mathcal{K}, \mathbf{a}_j) \qquad \text{for all } j \in \mathcal{J}. \end{aligned}$$

For an arbitrary $j \in \mathcal{J}$ we have that $\mathcal{F}(\mathcal{K}, \mathbf{a}_j)$ is an exposed face of $\mathcal{K}$ and because $\mathcal{F}$ is a maximal face we get that $\mathcal{F} = \mathcal{F}(\mathcal{K}, \mathbf{a}_j)$, completing the proof. $\square$

The converse is not true as if $\mathbf{a} \in \mathrm{Ext}(\mathcal{K}^*)$ then $\mathcal{F}(\mathcal{K}, \mathbf{a})$ is not necessarily a maximal face. We do however always get maximal faces from the exposed rays. Before we prove this we first need the following two trivial lemmas.

Lemma 8.16. *For $\mathbf{a} \in \mathcal{K}^*$ and $\mathbf{x} \in \mathcal{K}$,*

$$\mathbf{a} \in \mathcal{F}(\mathcal{K}^*, \mathbf{x}) \Leftrightarrow \langle \mathbf{a}, \mathbf{x} \rangle = 0 \Leftrightarrow \mathbf{x} \in \mathcal{F}(\mathcal{K}, \mathbf{a}).$$

Lemma 8.17. *For $\mathbf{a} \in \mathcal{K}^*$, $\lambda > 0$, we have that $\mathcal{F}(\mathcal{K}, \lambda\mathbf{a}) = \mathcal{F}(\mathcal{K}, \mathbf{a})$.*

Theorem 8.18. *If $\mathcal{K}$ is a proper cone and $\mathbf{a} \in \mathrm{Exp}(\mathcal{K}^*)$, then $\mathcal{F}(\mathcal{K}, \mathbf{a})$ is a maximal face of $\mathcal{K}$.*

Proof. Consider an arbitrary $\mathbf{a} \in \operatorname{Exp}(\mathcal{K}^*)$.
By the definition of an exposed ray and Theorem 8.13, there exists $\mathbf{x} \in \mathcal{K}$ such that $\mathcal{F}(\mathcal{K}^*, \mathbf{x}) = \{\alpha\mathbf{a} \mid \alpha \geq 0\}$.
From Lemma 8.16 this means that $\mathbf{x} \in \mathcal{F}(\mathcal{K}, \mathbf{a})$ and $\mathbf{x} \notin \mathcal{F}(\mathcal{K}, \mathbf{b})$ for all $\mathbf{b} \in \mathcal{K}^* \setminus \{\alpha\mathbf{a} \mid \alpha \geq 0\}$.
From Lemma 8.17 we have if $\mathbf{b} = \alpha\mathbf{a}$, where $\alpha > 0$, then $\mathcal{F}(\mathcal{K}, \mathbf{a}) = \mathcal{F}(\mathcal{K}, \mathbf{b})$.
Therefore there does not exist $\mathbf{b} \in \mathcal{K}^* \setminus \{\mathbf{0}\}$ such that $\mathcal{F}(\mathcal{K}, \mathbf{a}) \subset \mathcal{F}(\mathcal{K}, \mathbf{b})$.
This combined with Definition 8.5 and Theorems 8.4 and 8.13 gives the required result. □

8.2 Copositive & Completely Positive Cones of Order 2

We will now illustrate some of the theorems from the previous section with a quick example in Fig. 8.1. For this we use the copositive and completely positive cones in $\mathcal{S}^2$, which are proper cones and duals of each other. In order to show these in two dimensions we first use the svec operator to give an isomorphic mapping from $\mathcal{S}^2$ to $\mathbb{R}^3$.

$$\operatorname{svec}\left[\begin{pmatrix} x & y \\ y & z \end{pmatrix}\right] := \begin{pmatrix} x & \sqrt{2}y & z \end{pmatrix}^{\mathsf{T}},$$

which has the property

$$\langle A, B\rangle = \operatorname{trace}(AB) = \operatorname{svec}(A)^{\mathsf{T}} \operatorname{svec}(B).$$

We then consider the bases of these cones given by their intersections with the hyperplane $\boldsymbol{e}^{\mathsf{T}}\mathbf{x} = 1$.

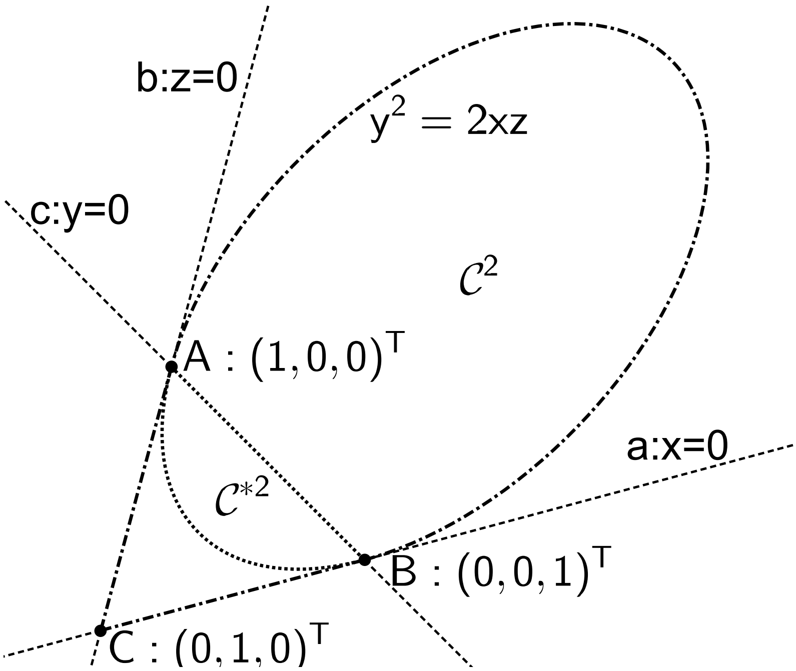
For these cones we have the following relationships between their extreme rays and their exposed faces:

i. A and B generate exposed rays of the completely positive cone, whilst the corresponding hyperplanes a and b give maximal faces of the copositive cone.

ii. A and B generate extreme but not exposed rays of the copositive cone, whilst the corresponding hyperplanes a and b give nonmaximal faces of the completely positive cone.

iii. C generates an exposed ray of the copositive cone, whilst the corresponding hyperplane c gives a maximal face of the completely positive cone.

Figure 8.1: The figure below is of bases of cones equivalent to the copositive and completely positive cones in $\mathcal{S}^2$, contained within "_ . _ . _" and "......" respectively. The equivalence between this figure and the cones is explained in Section 8.2. Letters in upper case label generators of rays and the equivalent letters in lower case label the corresponding hyperplanes.



8.3 Extreme Rays of the Copositive and Completely Positive Cones

In this section we will look at extreme rays of the copositive and completely positive cones. We will show that for $n \geq 2$, every extreme ray of the completely positive cone is also an exposed ray of it, and this is in contrast to the copositive cone for which we will give an example of extreme rays which are not exposed.

Theorem 8.19. *For $n \geq 2$, every extreme ray of the completely positive cone is also an exposed ray of it, i.e.*

$$\operatorname{Exp}(\mathcal{C}^{*n}) = \operatorname{Ext}(\mathcal{C}^{*n}) = \left\{ \mathbf{b}\mathbf{b}^{\mathsf{T}} \;\middle|\; \mathbf{b} \in \mathbb{R}^n_+ \setminus \{\mathbf{0}\} \right\}.$$

Proof. We begin with the following well-known result characterising the set of extreme rays of the completely positive cone:

$$\operatorname{Ext}(\mathcal{C}^{*}) = \left\{ \mathbf{b}\mathbf{b}^{\mathsf{T}} \;\middle|\; \mathbf{b} \in \mathbb{R}^n_+ \setminus \{\mathbf{0}\} \right\} \qquad \text{(see [BSM03, page 71]).}$$

For an arbitrary $\mathbf{b} \in \mathbb{R}^n_+ \setminus \{\mathbf{0}\}$, let $\{\mathbf{a}_1, \ldots, \mathbf{a}_{n-1}\} \subseteq \mathbb{R}^n$ be a set of linearly independent vectors which are perpendicular to $\mathbf{b}$. Now consider the exposed face of the completely positive cone given by $A = \sum_{j=1}^{n-1} \mathbf{a}_j \mathbf{a}_j^\mathsf{T} \in \mathcal{C}^n$,

$$\begin{aligned}
\mathcal{F}(\mathcal{C}^*, A) &= \{\textstyle\sum_i \mathbf{c}_i \mathbf{c}_i^\mathsf{T} \mid \mathbf{c}_i \in \mathbb{R}^n_+, \text{ for all } i, \quad 0 = \langle \sum_i \mathbf{c}_i \mathbf{c}_i^\mathsf{T}, \sum_j \mathbf{a}_j \mathbf{a}_j^\mathsf{T} \rangle\} \\
&= \{\textstyle\sum_i \mathbf{c}_i \mathbf{c}_i^\mathsf{T} \mid \mathbf{c}_i \in \mathbb{R}^n_+, \text{ for all } i, \quad 0 = \sum_{i,j} (\mathbf{c}_i^\mathsf{T} \mathbf{a}_j)^2\} \\
&= \{\textstyle\sum_i \mathbf{c}_i \mathbf{c}_i^\mathsf{T} \mid \mathbf{c}_i \in \mathbb{R}^n_+, \ \mathbf{c}_i^\mathsf{T} \mathbf{a}_j = 0 \text{ for all } i, j\} \\
&= \{\textstyle\sum_i \mathbf{c}_i \mathbf{c}_i^\mathsf{T} \mid \mathbf{c}_i = \alpha_i \mathbf{b}, \ \alpha_i \geq 0 \text{ for all } i\} \\
&= \{\alpha \mathbf{b}\mathbf{b}^\mathsf{T} \mid \alpha \geq 0\}.
\end{aligned}$$

Therefore the ray generated by $\mathbf{b}\mathbf{b}^\mathsf{T}$ is an exposed ray. □

For $n > 5$, finding the complete set of extreme rays of the copositive cone is still an open question. We do however have the following results for matrices which generate extreme rays.

Theorem 8.20. *For $n \geq 2$, we have the following results for the extreme rays of the copositive cone:*

i. $\alpha E_{ij} \in \operatorname{Ext}(\mathcal{C}^n)$, *where $i, j = 1, \ldots, n$, $\alpha > 0$, and this is all the nonnegative matrices which generate extreme rays of copositive cone.*

ii. $\mathbf{a}\mathbf{a}^\mathsf{T} \in \operatorname{Ext}(\mathcal{C})$, *where $\mathbf{a} \in \mathbb{R}^n \setminus (\mathbb{R}^n_+ \cup (-\mathbb{R}^n_+))$, and this, along with the relevant nonnegative matrices αE_{ii} from (i), is all the positive semidefinite matrices which generate extreme rays of copositive cone.*

iii. The set $\{X \in \operatorname{Ext}(\mathcal{C}) \mid (X)_{ij} \in \{-1, 0, +1\}, \ (X)_{ii} = +1 \text{ for all } i, j\}$, *was found in [HP73].*

iv. $M \in \operatorname{Ext}(\mathcal{C}) \Leftrightarrow PDMDP^\mathsf{T} \in \operatorname{Ext}(\mathcal{C})$, *where P is a permutation matrix and D is a diagonal matrix such that $(D)_{ii} > 0$ for all i.*

v. For $M \in \mathcal{C}^n \setminus \{0\}$, $B \in \mathbb{R}^{n \times m}$ *we have that* $\begin{pmatrix} M & B \\ B^\mathsf{T} & 0 \end{pmatrix} \in \operatorname{Ext}(\mathcal{C}^{n+m})$ *if and only if $B = 0$ and $M \in \operatorname{Ext}(\mathcal{C}^n)$.*

vi. If $\begin{pmatrix} M & \mathbf{m} \\ \mathbf{m}^\mathsf{T} & \mu \end{pmatrix} \in \operatorname{Ext}(\mathcal{C}^n) \setminus \mathcal{N}^n$, *then* $\begin{pmatrix} M & \mathbf{m} & \mathbf{m} \\ \mathbf{m}^\mathsf{T} & \mu & \mu \\ \mathbf{m}^\mathsf{T} & \mu & \mu \end{pmatrix} \in \operatorname{Ext}(\mathcal{C}^{n+1})$.

vii. For $\boldsymbol{\theta} \in \mathbb{R}^5$ we define the following matrices, which are referred to as the Hildebrand Matrices:

$$S(\boldsymbol{\theta}) := \begin{pmatrix}
1 & -\cos\theta_1 & \cos(\theta_1+\theta_2) & \cos(\theta_4+\theta_5) & -\cos\theta_5 \\
-\cos\theta_1 & 1 & -\cos\theta_2 & \cos(\theta_2+\theta_3) & \cos(\theta_5+\theta_1) \\
\cos(\theta_1+\theta_2) & -\cos\theta_2 & 1 & -\cos\theta_3 & \cos(\theta_3+\theta_4) \\
\cos(\theta_4+\theta_5) & \cos(\theta_2+\theta_3) & -\cos\theta_3 & 1 & -\cos\theta_4 \\
-\cos\theta_5 & \cos(\theta_5+\theta_1) & \cos(\theta_3+\theta_4) & -\cos\theta_4 & 1
\end{pmatrix},$$

i.e. $S(\boldsymbol{\theta}) \in \mathcal{S}^5$ *and for all* i *(considering modulus 5 on the indices) we have* $(S(\boldsymbol{\theta}))_{ii} = 1$, $(S(\boldsymbol{\theta}))_{i(i+1)} = -\cos\theta_i$, $(S(\boldsymbol{\theta}))_{i(i+2)} = \cos(\theta_i + \theta_{i+1})$.

We have that $S(\boldsymbol{\theta}) \in \operatorname{Ext}(\mathcal{C}^5)$ *for all* $\boldsymbol{\theta} \in \mathbb{R}^5_{++}$ *such that* $\mathbf{e}^\mathsf{T}\boldsymbol{\theta} < \pi$. *Furthermore, considering these matrices and* $S(\mathbf{0})$, *along with their transforms under the actions in* (iv), *we get all of the generators of extreme rays of* $\mathcal{C}^5$ *which are not in* $\mathcal{S}^5_+ + \mathcal{N}^5$. *The matrix* $S(\mathbf{0}) \in \operatorname{Ext}(\mathcal{C}^5)$, *is in fact referred to as the Horn matrix and has all entries equal to* ± 1, *so this case is already covered in* (iii).

Proof. Parts (i) and (ii) come directly from [HN63], whilst part (iii) comes from [HP73]. Part (iv) is trivial to show by transforming the coordinate basis. Part (v) follows trivially from [Dia62, Bau66]. Part (vi) comes directly from [Bau66]. Part (vii) is the subject of [Hil12]. □

Remark 8.21. A natural question that the reader may have about the Hildebrand matrices from the previous theorem is what happens when $\boldsymbol{e}^\mathsf{T}\boldsymbol{\theta} = \pi$, or when $\boldsymbol{\theta} \in \mathbb{R}^n_+ \setminus \mathbb{R}^n_{++}$ with $\boldsymbol{e}^\mathsf{T}\boldsymbol{\theta} < \pi$.

Considering the first case, when $\boldsymbol{e}^\mathsf{T}\boldsymbol{\theta} = \pi$, we have

$$S(\boldsymbol{\theta}) = \mathbf{a}\mathbf{a}^\mathsf{T} + \mathbf{b}\mathbf{b}^\mathsf{T},$$

where

$$\mathbf{a} = \begin{pmatrix} \cos\theta_1 \\ -1 \\ \cos\theta_2 \\ -\cos(\theta_2+\theta_3) \\ \cos(\theta_2+\theta_3+\theta_4) \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} \sin\theta_1 \\ 0 \\ -\sin\theta_2 \\ \sin(\theta_2+\theta_3) \\ -\sin(\theta_2+\theta_3+\theta_4) \end{pmatrix}.$$

Considering the second case, when $\boldsymbol{\theta} \in \mathbb{R}^n_+ \setminus \mathbb{R}^n_{++}$ with $\boldsymbol{e}^\mathsf{T}\boldsymbol{\theta} < \pi$, where without loss of generality we let $\theta_1 = 0$, it was noted in [DDGH13] that

$$S(\boldsymbol{\theta}) = \mathbf{c}\mathbf{c}^\mathsf{T} + \operatorname{Diag}(\mathbf{d}) S(\mathbf{0}) \operatorname{Diag}(\mathbf{d}),$$

where

$$\mathbf{c} = \begin{pmatrix} -\sin(\frac{1}{2}(\ \ \theta_2+\theta_3+\theta_4+\theta_5)) \\ \sin(\frac{1}{2}(\ \ \theta_2+\theta_3+\theta_4+\theta_5)) \\ -\sin(\frac{1}{2}(-\theta_2+\theta_3+\theta_4+\theta_5)) \\ \sin(\frac{1}{2}(-\theta_2-\theta_3+\theta_4+\theta_5)) \\ -\sin(\frac{1}{2}(-\theta_2-\theta_3-\theta_4+\theta_5)) \end{pmatrix},$$

$$\mathbf{d} = \begin{pmatrix} \cos(\frac{1}{2}(\ \ \theta_2+\theta_3+\theta_4+\theta_5)) \\ \cos(\frac{1}{2}(\ \ \theta_2+\theta_3+\theta_4+\theta_5)) \\ \cos(\frac{1}{2}(-\theta_2+\theta_3+\theta_4+\theta_5)) \\ \cos(\frac{1}{2}(-\theta_2-\theta_3+\theta_4+\theta_5)) \\ \cos(\frac{1}{2}(-\theta_2-\theta_3-\theta_4+\theta_5)) \end{pmatrix} \in \mathbb{R}^5_{++}.$$

We will now give an example of an extreme ray of the copositive cone which is not an exposed ray.

Theorem 8.22. *Let $n \geq 2$ and $i \in \{1,\dots,n\}$. Then E_{ii} generates a ray of the copositive cone $\mathcal{C}^n$ which is extreme but not exposed.*

Proof. From the previous theorem we have that $E_{ii} \in \mathrm{Ext}(\mathcal{C})$.
Suppose for the sake of contradiction that E_{ii} also generates an exposed ray of the copositive cone. This is only true if there exists an exposed face of the copositive cone which is equal to this ray. Therefore, by Theorem 8.13, there exists $B \in \mathcal{C}^*$ such that

$$\{\alpha E_{ii} \mid \alpha \geq 0\} = \mathcal{F}(\mathcal{C}, B) := \{A \in \mathcal{C} \mid \langle A, B\rangle = 0\}.$$

As $B \in \mathcal{C}^*$, we decompose it as $B = \sum_k \mathbf{b}_k \mathbf{b}_k^\mathsf{T}$, where $\mathbf{b}_k \in \mathbb{R}^n_+$ for all k.
As $E_{ii} = \boldsymbol{e}_i \boldsymbol{e}_i^\mathsf{T} \in \mathcal{F}(\mathcal{C}, B)$, we get $\langle \boldsymbol{e}_i \boldsymbol{e}_i^\mathsf{T}, B\rangle = 0$ and so $\mathbf{b}_k^\mathsf{T} \boldsymbol{e}_i = 0$ for all k.
We now consider the copositive matrix $E_{ij} = (\boldsymbol{e}_i \boldsymbol{e}_j^\mathsf{T} + \boldsymbol{e}_j \boldsymbol{e}_i^\mathsf{T})$, where $j \neq i$.

$$\langle E_{ij}, B\rangle = 2\textstyle\sum_k (\boldsymbol{e}_j^\mathsf{T} \mathbf{b}_k)(\mathbf{b}_k^\mathsf{T} \boldsymbol{e}_i) = 0.$$

Therefore $E_{ij} \in \mathcal{F}(\mathcal{C}, B) \setminus \{\alpha E_{ii} \mid \alpha \geq 0\} = \emptyset$, a contradiction. □

From the extension of Straszewicz's Theorem, we naturally have that the copositive cone does have exposed rays, and some of these are presented in the following theorem.

Theorem 8.23. *For $n \geq 2$, we have the following results for the exposed rays of the copositive cone:*

i. $\alpha E_{ij} \in \mathrm{Exp}(\mathcal{C}^n)$, *where* $i \neq j$, $\alpha > 0$,

ii. $\mathbf{a}\mathbf{a}^\mathsf{T} \in \mathrm{Exp}(\mathcal{C}^n)$, *where* $\mathbf{a} \in \mathbb{R}^n \setminus (\mathbb{R}^n_+ \cup (-\mathbb{R}^n_+))$,

iii. $M \in \mathrm{Exp}(\mathcal{C}^n)$ *for all* $M \in \mathrm{Ext}(\mathcal{C}^n)$ *such that* $(M)_{ij} = \pm 1$ *for all* i, j,

iv. $M \in \mathrm{Exp}(\mathcal{C}) \Leftrightarrow PDMDP^\mathsf{T} \in \mathrm{Exp}(\mathcal{C})$, *where P is a permutation matrix and D is a diagonal matrix such that $(D)_{ii} > 0$ for all i.*

v. $\widehat{M} = \begin{pmatrix} M & 0 \\ 0 & 0 \end{pmatrix} \in \mathrm{Exp}(\mathcal{C}^{n+m})$ *if and only if* $M \in \mathrm{Exp}(\mathcal{C}^n)$.

vi. If $\begin{pmatrix} M & \mathbf{m} \\ \mathbf{m}^\mathsf{T} & \mu \end{pmatrix} \in \operatorname{Exp}(\mathcal{C}^n) \setminus \mathcal{N}^n$*, then* $\begin{pmatrix} M & \mathbf{m} & \mathbf{m} \\ \mathbf{m}^\mathsf{T} & \mu & \mu \\ \mathbf{m}^\mathsf{T} & \mu & \mu \end{pmatrix} \in \operatorname{Exp}(\mathcal{C}^{n+1})$.

vii. Considering $S(\boldsymbol{\theta})$ *from Theorem 8.20vii, we have that* $S(\boldsymbol{\theta}) \in \operatorname{Exp}(\mathcal{C}^5)$ *for all* $\boldsymbol{\theta} \in \mathbb{R}^5_{++}$ *such that* $\mathbf{e}^\mathsf{T}\boldsymbol{\theta} < \pi$.

Proof. The proof of this involves going through each case separately. In each case we use Theorem 8.20 to show that the matrices generate extreme rays. We then construct a completely positive matrix A and show that the face $\mathcal{F}(\mathcal{C}^n, A)$ is equal to the extreme ray that we are interested in, and thus this is an exposed ray.

Going through each case separately in this thesis would involve several pages of repetitive workings, so instead we shall only prove part (*vii*). This then demonstrates the method described in the previous paragraph. For the proofs of parts (*i*) to (*vi*), the reader is pointed towards [Dic11, Theorem 4.6].

We consider an arbitrary $\boldsymbol{\theta} \in \mathbb{R}^5_{++}$ such that $\mathbf{e}^\mathsf{T}\boldsymbol{\theta} < \pi$. We note from Theorem 8.20 that $S(\boldsymbol{\theta}) \in \operatorname{Ext}(\mathcal{C}^n)$ and we define the nonnegative vectors

$$\mathbf{a}_1 = \begin{pmatrix} \sin\theta_2 \\ \sin(\theta_1+\theta_2) \\ \sin\theta_1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{a}_2 = \begin{pmatrix} 0 \\ \sin\theta_3 \\ \sin(\theta_2+\theta_3) \\ \sin\theta_2 \\ 0 \end{pmatrix}, \quad \mathbf{a}_3 = \begin{pmatrix} 0 \\ 0 \\ \sin\theta_4 \\ \sin(\theta_3+\theta_4) \\ \sin\theta_3 \end{pmatrix},$$

$$\mathbf{a}_4 = \begin{pmatrix} \sin\theta_4 \\ 0 \\ 0 \\ \sin\theta_5 \\ \sin(\theta_4+\theta_5) \end{pmatrix}, \quad \mathbf{a}_5 = \begin{pmatrix} \sin(\theta_5+\theta_1) \\ \sin\theta_5 \\ 0 \\ 0 \\ \sin\theta_1 \end{pmatrix}.$$

It was shown in [Hil12, Subsubsection 3.2.2] that

$$X \in \mathcal{C}^5, \quad \mathbf{a}_i^\mathsf{T} X \mathbf{a}_i = 0 \text{ for all } i \qquad \Leftrightarrow \qquad \exists \alpha \in \mathbb{R}_+ \text{ s.t. } X = \alpha S(\boldsymbol{\theta}).$$

Therefore, letting $A = \sum_{i=1}^5 \mathbf{a}_i \mathbf{a}_i^\mathsf{T} \in \mathcal{C}^{*n}$, we have that

$$\begin{aligned} \mathcal{F}\left(\mathcal{C}^5, A\right) &= \{X \in \mathcal{C}^5 \mid 0 = \langle X, A\rangle\} \\ &= \{X \in \mathcal{C}^5 \mid 0 = \textstyle\sum_i \mathbf{a}_i^\mathsf{T} X \mathbf{a}_i\} \\ &= \{X \in \mathcal{C}^5 \mid 0 = \mathbf{a}_i^\mathsf{T} X \mathbf{a}_i \text{ for all } i\} \\ &= \{\alpha S(\boldsymbol{\theta}) \mid \alpha \in \mathbb{R}_+\}. \end{aligned}$$

Therefore this ray is an exposed ray. □

8.4 Maximal Faces of the Copositive Cone

We can now use results developed in this chapter to give us the maximal faces of the copositive cone.

Lemma 8.24. *$\mathcal{F}$ is a maximal face of the copositive cone if and only if there exists $\mathbf{v} \in \mathbb{R}^n_+ \setminus \{\mathbf{0}\}$ such that*

$$\mathcal{F} = \mathcal{M}^n(\mathbf{v}) := \{X \in \mathcal{C}^n \mid \mathbf{v}^\mathsf{T} X \mathbf{v} = 0\}.$$

Proof. By Theorems 8.15, 8.18 and 8.19. □

We will now investigate the dimension of these faces. Without loss of generality we consider a vector defining a maximal face with the first p entries positive and the next $(n-p)$ entries equal to zero, for some $p \in \{1, \dots, n\}$. We can do this as for any nonzero vector the coordinate basis can easily be permuted so that this is so.

Lemma 8.25. *Let $\mathbf{v} = \begin{pmatrix} \widehat{\mathbf{v}} \\ 0 \end{pmatrix} \in \mathbb{R}^n_+$, where $\widehat{\mathbf{v}} \in \mathbb{R}^p_{++}$ and $p \in \{1, \dots, n\}$. Then*

$$\dim \mathcal{M}^n(\mathbf{v}) = \dim \mathcal{M}^p(\widehat{\mathbf{v}}) + \tfrac{1}{2}(n-p)(n+p+1).$$

Proof. In this proof we will subdivide the matrices as follows:

$$A = \begin{pmatrix} Y & W^\mathsf{T} \\ W & Z \end{pmatrix} \in \mathcal{S}^n$$

$$\begin{aligned} \text{such that} \quad & Y \in \mathcal{S}^p, \\ & W \in \mathbb{R}^{(n-p)\times p}, \\ & Z \in \mathcal{S}^{(n-p)}. \end{aligned}$$

If A is copositive then from Theorem 1.1 we get that Y and Z are copositive.

$$\begin{aligned} \mathcal{M}^n(\mathbf{v}) &= \{A \in \mathcal{C}^n \mid \mathbf{v}^\mathsf{T} A \mathbf{v} = 0\} \\ &= \left\{ A = \begin{pmatrix} Y & W^\mathsf{T} \\ W & Z \end{pmatrix} \in \mathcal{C}^n \;\middle|\; \widehat{\mathbf{v}}^\mathsf{T} Y \widehat{\mathbf{v}} = 0 \right\} \\ &= \left\{ A = \begin{pmatrix} Y & W^\mathsf{T} \\ W & Z \end{pmatrix} \in \mathcal{C}^n \;\middle|\; Y \in \mathcal{M}^p(\widehat{\mathbf{v}}) \right\}. \end{aligned}$$

To get an equation for the dimension of $\mathcal{M}^n(\mathbf{v})$, we will sandwich it between two other sets.

$$\mathcal{M}^n(\mathbf{v}) \subseteq \left\{ A = \begin{pmatrix} Y & W^\mathsf{T} \\ W & Z \end{pmatrix} \in \mathcal{S}^n \;\middle|\; \begin{matrix} Y \in \mathcal{M}^p(\widehat{\mathbf{v}}), \\ Z \in \mathcal{C}^{(n-p)}, \end{matrix} \quad W \in \mathbb{R}^{(n-p)\times p} \right\}$$

$$\begin{aligned}\text{Therefore,}\ \dim \mathcal{M}^n(\mathbf{v}) &\leq \dim \mathcal{M}^p(\widehat{\mathbf{v}}) + \dim \mathcal{C}^{(n-p)} + \dim \mathbb{R}^{(n-p)\times p} \\ &= \dim \mathcal{M}^p(\widehat{\mathbf{v}}) + \tfrac{1}{2}(n-p)(n-p+1) + (n-p)p.\end{aligned}$$

$$\mathcal{M}^n(\mathbf{v}) \supseteq \left\{ A = \begin{pmatrix} Y & W^\top \\ W & Z \end{pmatrix} \in \mathcal{S}^n \;\middle|\; \begin{matrix} Y \in \mathcal{M}^p(\widehat{\mathbf{v}}), \\ Z \in \mathcal{C}^{(n-p)}, \end{matrix} \quad W \in \mathbb{R}^{(n-p)\times p}_+ \right\}$$

$$\begin{aligned}\text{Therefore,}\ \dim \mathcal{M}^n(\mathbf{v}) &\geq \dim \mathcal{M}^p(\widehat{\mathbf{v}}) + \dim \mathcal{C}^{(n-p)} + \dim \mathbb{R}^{(n-p)\times p}_+ \\ &= \dim \mathcal{M}^p(\widehat{\mathbf{v}}) + \tfrac{1}{2}(n-p)(n-p+1) + (n-p)p. \ \square\end{aligned}$$

We now need the dimension of $\mathcal{M}^p(\widehat{\mathbf{v}})$, which we can find using the following lemma.

Lemma 8.26. *Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\} \subset \mathbb{R}^p_+ \setminus \{\mathbf{0}\}$ be a set of linearly independent vectors, where $\mathbf{u}_1 \in \mathbb{R}^p_{++}$ and $m < p$. Then we have that $\bigcap_{i=1}^m \mathcal{M}^p(\mathbf{u}_i)$ is an exposed face of the copositive cone which is isomorphic to the positive semidefinite cone $\mathcal{S}^{(p-m)}_+$.*

Proof. It is easy to see that the intersection of exposed faces is another exposed face. From the conditions given we have that

$$\begin{aligned}\bigcap_{i=1}^m \mathcal{M}^p(\mathbf{u}_i) &= \{A \in \mathcal{C}^p \mid \mathbf{u}_i^\top A \mathbf{u}_i = 0 \text{ for all } i = 1, \ldots, m\} \\ &= \{A \in \mathcal{S}^p_+ \mid \mathbf{u}_i^\top A \mathbf{u}_i = 0 \text{ for all } i = 1, \ldots, m\} \\ &\qquad \text{(from Theorem 1.1)} \\ &= \{\textstyle\sum_j \mathbf{a}_j \mathbf{a}_j^\top \mid \mathbf{a}_j \in \mathbb{R}^p,\ \mathbf{a}_j^\top \mathbf{u}_i = 0 \text{ for all } i, j\}.\end{aligned}$$

This set can now be seen to be isomorphic to $\mathcal{S}^{(p-m)}_+$. $\square$

We now use this to get the dimensions of the maximal faces.

Lemma 8.27. *Let $\mathbf{v} = \begin{pmatrix} \widehat{\mathbf{v}} \\ 0 \end{pmatrix} \in \mathbb{R}^n_+$, where $\widehat{\mathbf{v}} \in \mathbb{R}^p_{++}$ and $p \in \{1, \ldots, n\}$. Then*

$$\dim \mathcal{M}^n(\mathbf{v}) = \tfrac{1}{2}n(n+1) - p.$$

Proof. From Lemma 8.25 we have that

$$\dim \mathcal{M}^n(\mathbf{v}) = \dim \mathcal{M}^p(\widehat{\mathbf{v}}) + \tfrac{1}{2}(n-p)(n+p+1).$$

The previous lemma and $\widehat{\mathbf{v}} \in \mathbb{R}^p_{++}$ implies that $\mathcal{M}^p(\widehat{\mathbf{v}})$ is isomorphic to $\mathcal{S}^{p-1}_+$, so $\dim \mathcal{M}^p(\widehat{\mathbf{v}}) = \dim \mathcal{S}^{p-1}_+ = \frac{1}{2}p(p-1)$, completing the proof. $\square$

By considering permutations of the coordinate basis, we can now generalise the result from the previous lemma for all $\mathbf{v} \in \mathbb{R}^n_+ \setminus \{\mathbf{0}\}$ and combine this with Lemma 8.24 in order to give us the following theorem on the maximal faces of the copositive cone.

Theorem 8.28. *$\mathcal{F}$ is a maximal face of the copositive cone if and only if there exists $\mathbf{v} \in \mathbb{R}^n_+ \setminus \{\mathbf{0}\}$ such that*

$$\mathcal{F} = \mathcal{M}^n(\mathbf{v}) := \{X \in \mathcal{C}^n \mid \mathbf{v}^\mathsf{T} X \mathbf{v} = 0\}.$$

Furthermore, for a vector $\mathbf{v} \in \mathbb{R}^n_+ \setminus \{\mathbf{0}\}$ with p nonzero entries we have that

$$\dim \mathcal{M}^n(\mathbf{v}) = \tfrac{1}{2} n(n+1) - p.$$

An interesting corollary from this is that we get the following inequalities for the dimension of a maximal face $\mathcal{M}$ of $\mathcal{C}^n$,

$$\dim \mathcal{C}^{(n-1)} \leq \dim \mathcal{M} \leq \dim \mathcal{C}^n - 1.$$

Also note that for all $t \in \mathbb{Z}$ such that $\dim \mathcal{C}^{(n-1)} \leq t \leq \dim \mathcal{C}^n - 1$, there exists maximal face $\mathcal{M}$ of $\mathcal{C}^n$ such that $\dim \mathcal{M} = t$.

We can now show that the copositive cone has facets, as defined in Definition 8.1.

Theorem 8.29. *For $n \geq 2$, the copositive cone $\mathcal{C}^n$ has n facets and they are of the following form,*

$$\begin{aligned} \mathcal{M}^n(\boldsymbol{e}_i) &= \{A \in \mathcal{C}^n \mid (A)_{ii} = 0\} \qquad \text{for } i = 1, \ldots, n, \\ \mathcal{M}^n(\boldsymbol{e}_1) &= \left\{ \begin{pmatrix} 0 & b^\mathsf{T} \\ b & B \end{pmatrix} \,\middle|\, b \geq 0,\ B \in \mathcal{C}^{(n-1)} \right\}. \end{aligned}$$

(When being more specific about the form we took $i = 1$ for simplicity. The result can then be extended by permuting the coordinate basis.)

Proof. From Theorem 8.28, it can be clearly seen that the facets of the copositive cone are produced by vectors with only one nonzero entry. It can also be clearly seen that multiplying a vector by a strictly positive constant does not change the face that it describes, therefore all the facets can be produced by the unit vectors $\boldsymbol{e}_i$ for $i = 1, \ldots, n$. Using this we now get the following as the facets.

$$\begin{aligned} \mathcal{M}^n(\boldsymbol{e}_i) &= \{A \in \mathcal{C}^n \mid \boldsymbol{e}_i^\mathsf{T} A \boldsymbol{e}_i = 0\} \\ &= \{A \in \mathcal{C}^n \mid (A)_{ii} = 0\}, \\ \dim \mathcal{M}^n(\boldsymbol{e}_i) &= \tfrac{1}{2} n(n+1) - 1 \\ &= \dim \mathcal{C}^n - 1. \end{aligned}$$

In order to be more specific about the form that the facets take we first note that the conditions we give in the form are trivially sufficient for the matrix being on the face. Using Theorem 1.1 we see that these conditions are also necessary. □

8.5 Maximal Faces of the Completely Positive Cone

We were able to find all the maximal faces of the copositive cone due to the fact that we know all of the extreme rays of the completely positive cone and all of these rays are also exposed rays. Unfortunately, from Theorem 8.22 we have that the copositive cone has extreme rays which are not exposed, and when $n > 5$, finding all of the extreme rays of $\mathcal{C}^n$ is still an open question. We can however consider some of the extreme rays which we do know. In particular, by Theorem 8.18, the exposed rays in Theorem 8.23 give maximal faces of the completely positive cone. In [DA13] the authors consider the exposed face of the completely positive cone given by the Horn matrix and show that the dimension of this face is equal to 10. As this matrix is in the set $\{X \in \operatorname{Ext}(\mathcal{C}^5) \mid (X)_{ij} = \pm 1 \text{ for all } i\}$, we now see that the face is a maximal face. In this section we will look at some more maximal faces of the completely positive cone, although we start by presenting the following two lemmas for general faces of the completely positive cone.

Lemma 8.30. *Let M be a copositive matrix, P be a permutation matrix and D be a diagonal matrix such that $(D)_{ii} > 0$ for all i, then*

$$\dim(\mathcal{F}\left(\mathcal{C}^*, PDMDP^\mathsf{T}\right)) = \dim(\mathcal{F}\left(\mathcal{C}^*, M\right)).$$

Proof. This is trivial to prove by transforming the coordinate basis. □

Lemma 8.31. *Let $\emptyset \neq \mathcal{I} \subseteq \{1, \ldots, n\}$ and let $X \in \mathcal{C}^n$ be such that $(X)_{ii} > 0$ for all $i \in \mathcal{I}$. Then*

$$\dim(\mathcal{F}\left(\mathcal{C}^{*n}, X\right)) \leq \tfrac{1}{2}n(n+1) - |\mathcal{I}| = \dim \mathcal{C}^{*n} - |\mathcal{I}|.$$

Proof. In this proof, for simplicity of notation, we let $\dim(\mathcal{F}\left(\mathcal{C}^{*n}, X\right)) = m$.

If we consider the diagonal matrix $D_{\pm\delta,i} := I \pm \delta E_{ii}$ for $0 < \delta < 1$ then from Theorem 1.1 we have that the following matrix is copositive,

$$D_{\pm\delta,i} X D_{\pm\delta,i} = X \pm \delta \widetilde{X}_i + \delta^2 \widehat{X}_i,$$

where $\widetilde{X}_i := E_{ii}X + XE_{ii}$ and $\widehat{X}_i := (X)_{ii}E_{ii}$.

As $(X)_{ii} > 0$ for all $i \in \mathcal{I}$, we have that $\{\widetilde{X}_i \mid i \in \mathcal{I}\}$ is a set of $|\mathcal{I}|$ linearly independent symmetric matrices.

As $\mathbf{0} \in \mathcal{F}(\mathcal{C}^*, X)$ and $\dim \mathcal{F}(\mathcal{C}^*, X) = m$, there exists a set of linearly independent symmetric matrices $\{A_1, \ldots, A_m\} \subseteq \mathcal{F}(\mathcal{C}^*, X)$. We have that $\langle A_j, X\rangle = 0$ for all j.

For all $i \in \mathcal{I}$, for all $j = 1, \ldots, m$ and for all $\delta \in (0, 1)$, we have $A_j \in \mathcal{C}^*$ and $D_{\pm\delta,i} X D_{\pm\delta,i} \in \mathcal{C}$ and therefore, by the definition of the dual,

$$\begin{aligned} 0 &\leq \frac{1}{\delta}\langle A_j, D_{\pm\delta,i} X D_{\pm\delta,i}\rangle \\ &= \pm\langle A_j, \widetilde{X}_i\rangle + \delta\langle A_j, \widehat{X}_i\rangle \end{aligned}$$

Letting $\delta \to 0$ we get that $\langle A_j, \widetilde{X}_i\rangle = 0$ for all $i \in \mathcal{I}$, $j = 1, \ldots, m$.
Therefore $\{\widetilde{X}_i \mid i \in \mathcal{I}\} \cup \{A_1, \ldots, A_m\}$ is a set of linearly independent symmetric matrices, and this implies that $|\mathcal{I}| + m \leq \dim \mathcal{S}^n = \frac{1}{2}n(n+1)$. ☐

We next consider the following theorems on some maximal and nonmaximal faces of the completely positive cone.

Theorem 8.32. *For $n \geq 2$ and $i, j \in \{1, \ldots, n\}$ such that $i \neq j$, we have*

$$\begin{aligned} \mathcal{F}(\mathcal{C}^{*n}, E_{ij}) &= \{B \in \mathcal{C}^{*n} \mid (B)_{ij} = 0\}, \\ &= \{\textstyle\sum_k \mathbf{c}_k \mathbf{c}_k^\mathsf{T} \mid \mathbf{c}_k \in \mathbb{R}^n_+,\ (\mathbf{c}_k)_i (\mathbf{c}_k)_j = 0 \text{ for all } k\}, \\ \mathcal{F}(\mathcal{C}^{*n}, E_{ii}) &= \{B \in \mathcal{C}^{*n} \mid (B)_{il} = 0 \text{ for all } l = 1, \ldots, n\} \\ &\subset \mathcal{F}(\mathcal{C}^{*n}, E_{ij}). \end{aligned}$$

Proof. These results follow trivially from considering decompositions of a completely positive matrix. ☐

Theorem 8.33. *For $n \geq 2$, we have the following results on maximal faces of the completely positive cone:*

i. *For $\mathbf{a} \in \mathbb{R}^n \setminus (\mathbb{R}^n_+ \cup (-\mathbb{R}^n_+))$, we have that $\mathcal{F}\left(\mathcal{C}^{*n}, \mathbf{a}\mathbf{a}^\mathsf{T}\right)$ is a maximal face of the completely positive cone with dimension equal to $\frac{1}{2}n(n-1)$.*

ii. *For $M \in \mathrm{Ext}(\mathcal{C}^n)$ such that $(M)_{ij} = \pm 1$ for all i, j, we have that $\mathcal{F}(\mathcal{C}^{*n}, M)$ is a maximal face of the completely positive cone with dimension equal to $\frac{1}{2}n(n-1)$.*

iii. *For $i \neq j$ we have that $\mathcal{F}(\mathcal{C}^{*n}, E_{ij})$ is a maximal face of the completely positive cone with dimension equal to $\frac{1}{2}n(n+1) - 1$, i.e. they are facets. These are in fact the only facets of the completely positive cone, and thus the completely positive cone has $\frac{1}{2}n(n-1)$ facets. The form of these facets is given in Theorem 8.32.*

Proof. We shall simply give a sketch of the proof of this, referring the reader to [Dic11, Section 6] for a full proof.

From Theorems 8.18 and 8.23, we get that the faces considered are indeed maximal faces. We next find lower and upper bounds on the dimension of these faces, with the lower and upper bounds for each face being equal. Lower bounds are found through the construction of linearly independent completely positive matrices contained in the faces, whilst upper bounds are found in the following ways:

i. It can be seen that $\mathcal{F}\left(\mathcal{C}^{*n}, \mathbf{a}\mathbf{a}^\mathsf{T}\right) \subseteq \mathcal{F}\left(\mathcal{S}_+^n, \mathbf{a}\mathbf{a}^\mathsf{T}\right)$ and thus we have that $\dim \mathcal{F}\left(\mathcal{C}^{*n}, \mathbf{a}\mathbf{a}^\mathsf{T}\right) \leq \dim \mathcal{F}\left(\mathcal{S}_+^n, \mathbf{a}\mathbf{a}^\mathsf{T}\right) = \frac{1}{2}n(n-1)$.

ii. From Lemma 8.31 we have that $\dim \mathcal{F}\left(\mathcal{C}^{*n}, M\right) \leq \frac{1}{2}n(n+1) - n$.

iii. All exposed faces of a cone in $\mathcal{S}^n$ have dimension strictly less than $\dim(\mathcal{S}^n) = \frac{1}{2}n(n+1)$.

We are then left to prove that $\mathcal{F}\left(\mathcal{C}^{*n}, E_{ij}\right)$ are the only facets, which comes from Lemmas 8.14 and 8.31, and Theorems 1.1 and 8.32. $\square$

8.6 Lower Bound on Dimension of Maximal Faces of the Completely Positive Cone

In all of our examples of maximal faces of the completely positive cone so far looked at, we have had that their dimensions were greater than or equal to $\frac{1}{2}n(n-1)$. From this, a natural question is whether this is in fact a lower bound on the dimension of the maximal faces for the completely positive cone, as it was for the copositive cone. For $n \leq 4$ we know all the extreme rays of the copositive cone and from the analysis in the previous section we see that all maximal faces have dimension equal to either $\frac{1}{2}n(n-1)$ or $\frac{1}{2}n(n+1)-1$, and thus the answer to the question is yes for $n \leq 4$. However, in this section we shall show that even for $n = 5$ the answer to this question is no. In order to do this we first need the following theorem.

Theorem 8.34. *Consider a matrix $A \in \mathcal{C} \setminus \{0\}$. Let $\{\mathcal{X}_1, \ldots, \mathcal{X}_m\}$ be such that $\mathcal{X}_i \subset \mathbb{R}_+^n$ is a finite set for all i and*

$$\mathcal{V}^A = \bigcup_{i=1}^{m} \operatorname{conic} \mathcal{X}_i.$$

Using Method 6.5 we can always find such a set. We now define the following *for $i = 1, \ldots, m$:*

$$\begin{aligned}\mathcal{Y}_i &= \{\mathbf{b}_1 + \mathbf{b}_2 \mid \mathbf{b}_1, \mathbf{b}_2 \in \mathcal{X}_i\},\\ \mathcal{Z}_i &= \{\mathbf{b}\mathbf{b}^\mathsf{T} \mid \mathbf{b} \in \mathcal{Y}_i\}\\ &= \{(\mathbf{b}_1 + \mathbf{b}_2)(\mathbf{b}_1 + \mathbf{b}_2)^\mathsf{T} \mid \mathbf{b}_1, \mathbf{b}_2 \in \mathcal{X}_i\}.\end{aligned}$$

Then we have that $\operatorname{cone}\left(\bigcup_{i=1}^m \mathcal{Z}_i\right) \subseteq \mathcal{F}(\mathcal{C}^*, A) \subseteq \operatorname{span}\left(\bigcup_{i=1}^m \mathcal{Z}_i\right)$ *and thus* $\dim \mathcal{F}(\mathcal{C}^*, A) = \dim\left(\operatorname{span}\left(\bigcup_{i=1}^m \mathcal{Z}_i\right)\right)$*. We also note that as $\mathcal{Z}_i$ is a finite set, this value is relatively easy to compute.*

Proof. This comes directly from noting that

$$\begin{aligned}\mathcal{F}(\mathcal{C}^*, A) &= \{\textstyle\sum_i \mathbf{b}_i\mathbf{b}_i^\mathsf{T} \mid \mathbf{b}_i \in \mathbb{R}^n_+ \text{ for all } i,\ \sum_i \mathbf{b}_i^\mathsf{T} A\mathbf{b}_i = 0\}\\ &= \operatorname{conic}\{\mathbf{b}\mathbf{b}^\mathsf{T} \mid \mathbf{b} \in \mathcal{V}^A\}\\ &\supseteq \operatorname{cone}\left(\bigcup_{i=1}^m \mathcal{Z}_i\right),\end{aligned}$$

and

$$\begin{aligned}&\mathcal{F}(\mathcal{C}^*, A)\\ &= \operatorname{conic}\left\{\mathbf{b}\mathbf{b}^\mathsf{T} \,\middle|\, \mathbf{b} \in \mathcal{V}^A\right\}\\ &\subseteq \operatorname{span} \bigcup_{i=1}^m \left\{\mathbf{b}\mathbf{b}^\mathsf{T} \,\middle|\, \mathbf{b} \in \operatorname{conic} \mathcal{X}_i\right\}\\ &= \operatorname{span} \bigcup_{i=1}^m \left\{(\textstyle\sum_j \alpha_j \mathbf{b}_j)(\sum_k \alpha_k \mathbf{b}_k)^\mathsf{T} \,\middle|\, \alpha_l \geq 0,\ \mathbf{b}_l \in \mathcal{X}_i \text{ for all } l\right\}\\ &= \operatorname{span} \bigcup_{i=1}^m \left\{\begin{array}{r}\frac{1}{2}\sum_{j,k} \alpha_j\alpha_k(\mathbf{b}_j + \mathbf{b}_k)(\mathbf{b}_j + \mathbf{b}_k)^\mathsf{T}\\ - \frac{1}{4}\sum_{j,k} \alpha_j\alpha_k(2\mathbf{b}_j)(2\mathbf{b}_j)^\mathsf{T}\end{array} \,\middle|\, \alpha_l \geq 0,\ \mathbf{b}_l \in \mathcal{X}_i \text{ for all } l\right\}\\ &\subseteq \operatorname{span}\left(\bigcup_{i=1}^m (\operatorname{span} \mathcal{Z}_i)\right)\\ &= \operatorname{span}\left(\bigcup_{i=1}^m \mathcal{Z}_i\right).\end{aligned}$$

□

We now consider the completely positive cone of order five.

Theorem 8.35. *The following are all of the maximal faces of the completely positive cone of order five, where we consider $S(\boldsymbol{\theta})$ from Theorem 8.20vii:*

i. $\mathcal{F}\left(\mathcal{C}^{*5}, E_{ij}\right)$*, where $i \neq j$. These faces have dimension equal to 14.*

ii. $\mathcal{F}\left(\mathcal{C}^{*5}, \mathbf{a}\mathbf{a}^{\mathsf{T}}\right)$, *where* $\mathbf{a} \in \mathbb{R}^5 \setminus (\mathbb{R}^n_+ \cup (-\mathbb{R}^n_+))$. *These faces have dimension equal to 10,*

iii. $\mathcal{F}\left(\mathcal{C}^{*5}, PDS(\mathbf{0})DP^{\mathsf{T}}\right)$, *where* P *is a permutation matrix and* D *is a diagonal matrix such that* $(D)_{ii} > 0$ *for all* i. *These faces have dimension equal to 10,*

iv. $\mathcal{F}\left(\mathcal{C}^{*5}, PDS(\boldsymbol{\theta})DP^{\mathsf{T}}\right)$, *where* $\boldsymbol{\theta} \in \mathbb{R}^5_{++}$ *such that* $\boldsymbol{e}^{\mathsf{T}}\boldsymbol{\theta} < \pi$ *and* P *is a permutation matrix and* D *is a diagonal matrix such that* $(D)_{ii} > 0$ *for all* i. *These faces have dimension equal to 5.*

Proof. All of the results in this theorem, except for the dimension of the faces $\mathcal{F}\left(\mathcal{C}^{*5}, PDS(\boldsymbol{\theta})DP^{\mathsf{T}}\right)$, come directly from Theorems 8.15, 8.18 and 8.23 and the results in Section 8.5.

We consider an arbitrary $\boldsymbol{\theta} \in \mathbb{R}^5_{++}$ such that $\boldsymbol{e}^{\mathsf{T}}\boldsymbol{\theta} < \pi$ and consider the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_5 \in \mathbb{R}^5_+$ from the proof of Theorem 8.23. It was shown in [Hil12] that $\mathcal{V}^{S(\boldsymbol{\theta})} = \bigcup_{i=1}^5 \operatorname{cone}\{\mathbf{a}_i\}$, and thus from Lemma 8.30 and Theorem 8.34 we get the required result. □

This can now be seen to be a valid counter example to the statement in our question by noting that for $n = 5$ we have $\frac{1}{2}n(n-1) = 10$.

Part III

Approximations

In Section 1.3 and Chapter 3, we looked at the $\mathcal{NP}$-hardness of copositive optimisation. From this, we would not expect there to be efficient exact algorithms for solving general copositive optimisation problems. Instead, we consider replacing the copositive cone with approximation hierarchies.

For sequences of convex cones, $\{\mathcal{I}_r \mid r \in \mathbb{Z}_+\}$ and $\{\mathcal{O}_r \mid r \in \mathbb{Z}_+\}$, such that $\mathcal{I}_r, \mathcal{O}_r \subseteq \mathcal{S}^n$ for all r, we say that $\mathcal{I}_r$ and $\mathcal{O}_r$ are respectively inner and outer approximation hierarchies for a closed convex set $\mathcal{K} \subseteq \mathcal{S}^n$ if

$$\mathcal{I}_0 \subseteq \mathcal{I}_1 \subseteq \ldots \subseteq \bigcup_{r\in\mathbb{Z}_+} \mathcal{I}_r \subseteq \mathcal{K}, \tag{8.1}$$

$$\mathcal{O}_0 \supseteq \mathcal{O}_1 \supseteq \ldots \supseteq \bigcap_{r\in\mathbb{Z}_+} \mathcal{O}_r \supseteq \mathcal{K}. \tag{8.2}$$

We say that an inner approximation hierarchy converges if

$$\operatorname{cl}\left(\bigcup_{r\in\mathbb{Z}_+} \mathcal{I}_r\right) = \mathcal{K}. \tag{8.3}$$

In other words, for all $X \in \operatorname{int}\mathcal{K}$, there exists an $r \in \mathbb{Z}_+$ such that $X \in \mathcal{I}_r$.

Similarly, we say that an outer approximation hierarchy converges if

$$\bigcap_{r\in\mathbb{Z}_+} \mathcal{O}_r = \mathcal{K}. \tag{8.4}$$

In other words, for all $X \notin \mathcal{K}$, there exists an $r \in \mathbb{Z}_+$ such that $X \notin \mathcal{O}_r$.

Note that by standard results on duality (see Section 1.2.3), we have that (8.1), (8.2), (8.3) and (8.4) are respectively equivalent to

$$\mathcal{I}_0^* \supseteq \mathcal{I}_1^* \supseteq \ldots \supseteq \bigcap_{r\in\mathbb{Z}_+} \mathcal{I}_r^* \supseteq \mathcal{K}^*,$$

$$\mathcal{O}_0^* \subseteq \mathcal{O}_1^* \subseteq \ldots \subseteq \bigcup_{r\in\mathbb{Z}_+} \mathcal{O}_r^* \subseteq \mathcal{K}^*,$$

$$\bigcap_{r\in\mathbb{Z}_+} \mathcal{I}_r^* = \mathcal{K}^*,$$

$$\operatorname{cl}\left(\bigcup_{r\in\mathbb{Z}_+} \mathcal{O}_r^*\right) = \mathcal{K}^*.$$

We shall now consider convergent approximation hierarchies being used for optimisation. We consider the following arbitrary optimisation problems, where $\{\mathcal{I}_r \mid r \in \mathbb{Z}_+\}$ and $\{\mathcal{O}_r \mid r \in \mathbb{Z}_+\}$ are respectively convergent inner and outer approximation hierarchies for a closed convex cone $\mathcal{K} \subseteq \mathcal{S}^n$, and $A_0, \ldots, A_m \in \mathcal{S}^n$, and $b_1, \ldots, b_m \in \mathbb{R}$.

$$\begin{aligned} \min_{X \in \mathcal{S}^n} \quad & \langle A_0, X \rangle \\ \text{s.t.} \quad & \langle A_i, X \rangle = b_i \qquad \text{for all } i = 1, \ldots, m \\ & X \in \mathcal{K}, \end{aligned} \tag{8.5}$$

$$\begin{aligned} \min_{X \in \mathcal{S}^n} \quad & \langle A_0, X \rangle \\ \text{s.t.} \quad & \langle A_i, X \rangle = b_i \qquad \text{for all } i = 1, \ldots, m \\ & X \in \mathcal{I}_r, \end{aligned} \tag{8.6$_r$}$$

$$\begin{aligned} \min_{X \in \mathcal{S}^n} \quad & \langle A_0, X \rangle \\ \text{s.t.} \quad & \langle A_i, X \rangle = b_i \qquad \text{for all } i = 1, \ldots, m \\ & X \in \mathcal{O}_r. \end{aligned} \tag{8.7$_r$}$$

For all $r \in \mathbb{Z}_+$, we have that

$$\text{Feas}\,(8.6_r) \subseteq \text{Feas}(8.6_{r+1}) \subseteq \text{Feas}\,(8.5) \subseteq \text{Feas}(8.7_{r+1}) \subseteq \text{Feas}\,(8.7_r),$$

and thus

$$\text{Val}\,(8.6_r) \geq \text{Val}(8.6_{r+1}) \geq \text{Val}\,(8.5) \geq \text{Val}(8.7_{r+1}) \geq \text{Val}\,(8.7_r).$$

Furthermore, if (8.5) is strictly feasible, then it can be seen that

$$\lim_{r \to \infty} \text{Val}\,(8.6_r) = \text{Val}\,(8.5).$$

Similarly, if the dual problem to (8.5) is strictly feasible, then it can be seen that

$$\lim_{r \to \infty} \text{Val}\,(8.7_r) = \text{Val}\,(8.5).$$

In the following chapters, we shall look at the most commonly used approximation hierarchies for the copositive cone. In fact, we shall often consider approximations to generalisations of the copositive cone, which will allow for improved understanding of the hierarchies.

Chapter 9

Simplicial Partitioning*

9.1 Introduction

In this chapter we will consider simplicial partitioning, which has shown itself to not just be useful in copositive optimisation, but also more generally in the field of Nonlinear Optimisation [BE12, BD08, Hor76, Hor97, Kea78, Tuy91a, TH88].

By *simplices*, we mean nonempty sets of the form $\Delta = \operatorname{conv}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$, where $\mathbf{v}_1, \ldots, \mathbf{v}_p \in \mathbb{R}^n$ are affinely independent (equivalently $\dim \Delta = p - 1$). Note that if $p = 3$, then Δ is a triangle. We let $V(\Delta) := \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ be the set of *vertices* of Δ and let $E(\Delta) := \{\{\mathbf{v}_i, \mathbf{v}_j\} \mid i, j = 1, \ldots, p : \ i < j\}$ be the set of *edges* of Δ.

We define the *diameter* of a simplex as follows,

$$d(\Delta) := \max\{\|\mathbf{x} - \mathbf{y}\|_2 \mid \mathbf{x}, \mathbf{y} \in \Delta\} = \max\{\|\mathbf{u} - \mathbf{v}\|_2 \mid \mathbf{u}, \mathbf{v} \in V(\Delta)\}.$$

We say that a sequence of simplices $\{\Delta_i \mid i \in \mathbb{Z}_+\}$ is a sequence of nested simplices if $\Delta_{i+1} \subseteq \Delta_i$ for all i. We then naturally have that $d(\Delta_{i+1}) \leq d(\Delta_i)$ for all i. We say that the sequence is exhaustive if $\lim_{i\to\infty} d(\Delta_i) = 0$.

If we have a simplex, Δ, and a finite set of simplices, $\mathcal{P} = \{\Delta_i \mid i \in \mathcal{I}\}$, such that

$$\begin{aligned} \Delta &= \bigcup_{i\in\mathcal{I}} \Delta_i, \\ \Delta_i \cap \Delta_j &= \operatorname{rbd} \Delta_i \cap \operatorname{rbd} \Delta_j \qquad \text{for all } i, j \in \mathcal{I} : \ i \neq j, \end{aligned}$$

then we say that $\mathcal{P}$ is a *simplicial partition* of Δ. In this thesis, the only partitions that we consider are simplicial partitions and we shall thus simply

---

*Submitted as:

[Dic13] P.J.C. Dickinson. On the exhaustivity of simplicial partitioning. *Journal of Global Optimization*, in print. DOI: 10.1007/s10898-013-0040-7 refer to them as partitions. We now note the following well-known result on partitions.

Lemma 9.1. *Consider a partition* $\mathcal{P} = \{\Delta_i \mid i \in \mathcal{I}\}$ *of a simplex* Δ*, as defined above. Then we have that* $\dim \Delta_i = \dim \Delta$ *for all* $i \in \mathcal{I}$*.*

Proof. We suppose, for the sake of contradiction, that there exists an $i \in \mathcal{I}$ such that $0 \leq \dim \Delta_i < \dim \Delta$, and we let $\mathbf{x} \in \Delta_i \setminus \operatorname{rbd} \Delta_i \neq \emptyset$. Due to the fact that $\dim \Delta_i < \dim \Delta$, and Δ being closed and convex, we have $\Delta = \operatorname{cl}(\Delta \setminus \Delta_i) \subseteq \bigcup_{j \in \mathcal{I} \setminus \{i\}} \Delta_j$. Therefore there exists a $j \in \mathcal{I} \setminus \{i\}$ such that $\mathbf{x} \in \Delta_j$, and we have that $\mathbf{x} \in (\Delta_i \cap \Delta_j) \setminus (\operatorname{rbd} \Delta_i \cap \operatorname{rbd} \Delta_j)$, which is a contradiction. $\square$

For a partition, $\mathcal{P}$, we define its set of vertices, its set of edges and its diameter as follows respectively:

$$V(\mathcal{P}) := \bigcup_{\Delta_i \in \mathcal{P}} V(\Delta_i), \qquad E(\mathcal{P}) := \bigcup_{\Delta_i \in \mathcal{P}} E(\Delta_i),$$

$$d(\mathcal{P}) := \max\{d(\Delta_i) \mid \Delta_i \in \mathcal{I}\} = \max\{\|\mathbf{u} - \mathbf{v}\|_2 \mid \{\mathbf{u}, \mathbf{v}\} \in E(\mathcal{P})\}.$$

For two partitions $\mathcal{P}_1 = \{\Delta_i \mid i \in \mathcal{I}\}$ and $\mathcal{P}_2 = \{\Delta_j \mid j \in \mathcal{J}\}$, we say that $\mathcal{P}_1$ is nested in $\mathcal{P}_2$ if for all $i \in \mathcal{I}$ there exists $j \in \mathcal{J}$ such that $\Delta_i \subseteq \Delta_j$. Naturally, we then have that $d(\mathcal{P}_1) \leq d(\mathcal{P}_2)$. We say that a sequence of partitions $\{\mathcal{P}_k \mid k \in \mathbb{Z}_+\}$ is a sequence of nested partitions if $\mathcal{P}_{k+1}$ is nested in $\mathcal{P}_k$ for all $k \in \mathbb{Z}_+$.

In this chapter, we consider sequences of nested partitions $\{\mathcal{P}_k \mid k \in \mathbb{Z}_+\}$ such that $\mathcal{P}_0 = \{\Delta\}$ and $\Delta = \bigcup_{\Delta_i \in \mathcal{P}_k} \Delta_i$ for all $k \in \mathbb{Z}_+$. It is then desirable that $\lim_{k \to \infty} d(\mathcal{P}_k) = 0$, and when this holds we say that the sequence is *exhaustive*. In the following subsection we look at how this can be used to provide approximation hierarchies for the copositive cone. In Section 9.1.2, we consider three possible methods for going from one partition to the next one in the sequence. Finally, in the remaining sections of this chapter, we look at when we can guarantee that the sequence of partitions is exhaustive, or give a counter-example to this happening.

9.1.1 Approximation hierarchy

In this subsection we shall review results from [BD08, BD09] on how simplicial partitions can provide approximation hierarchies for the copositive cone.

We define the standard simplex as follows,

$$\Delta^S := \{\mathbf{x} \in \mathbb{R}^n_+ \mid \boldsymbol{e}^\mathsf{T} \mathbf{x} = 1\}.$$

It can then be seen that

$$\mathcal{C}^n = \{X \in \mathcal{S}^n \mid \mathbf{v}^\mathsf{T} X \mathbf{v} \geq 0 \text{ for all } \mathbf{v} \in \Delta^S\},$$
$$\mathrm{int}(\mathcal{C}^n) = \{X \in \mathcal{S}^n \mid \mathbf{v}^\mathsf{T} X \mathbf{v} > 0 \text{ for all } \mathbf{v} \in \Delta^S\}.$$

Now, as in [BD09], for an arbitrary partition $\mathcal{P}$ of Δ^S, we define $\mathcal{O}_\mathcal{P}, \mathcal{I}_\mathcal{P} \subset \mathcal{S}^n$ as follows:

$$\mathcal{O}_\mathcal{P} := \{X \in \mathcal{S}^n \mid \mathbf{u}^\mathsf{T} X \mathbf{u} \geq 0 \text{ for all } \mathbf{u} \in V(\mathcal{P})\},$$
$$\mathcal{I}_\mathcal{P} := \{X \in \mathcal{O}_\mathcal{P} \mid \mathbf{u}^\mathsf{T} X \mathbf{v} \geq 0 \text{ for all } \{\mathbf{u}, \mathbf{v}\} \in E(\mathcal{P})\}.$$

From the following two lemmas we see that if we have an exhaustive sequence of nested partitions, then the sets above provide convergent inner and outer approximation hierarchies to the copositive cone.

Lemma 9.2. *Consider two partitions $\mathcal{P}, \mathcal{Q}$ of Δ^S such that $\mathcal{P}$ is nested in $\mathcal{Q}$. Then we have that $\mathcal{O}_\mathcal{P} \subseteq \mathcal{O}_\mathcal{Q}$ and $\mathcal{I}_\mathcal{P} \supseteq \mathcal{I}_\mathcal{Q}$.*

Proof. As $\mathcal{P}, \mathcal{Q}$ are both partitions of the same set and $\mathcal{P}$ is nested in $\mathcal{Q}$, it can be seen that $V(\mathcal{Q}) \subseteq V(\mathcal{P})$. Therefore $\mathcal{O}_\mathcal{P} \subseteq \mathcal{O}_\mathcal{Q}$.

We now consider an arbitrary $X \in \mathcal{I}_\mathcal{Q}$ and arbitrary vertices $\mathbf{u}, \mathbf{v}$ of $\mathcal{P}$ such that either $\mathbf{u} = \mathbf{v}$ or $\{\mathbf{u}, \mathbf{v}\}$ is an edge of $\mathcal{P}$. There exists a subsimplex $\Delta = \mathrm{conv}\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ of $\mathcal{Q}$ such that $\mathbf{u}, \mathbf{v} \in \Delta$. Therefore, there exists $\boldsymbol{\theta}, \boldsymbol{\varphi} \in \Delta^S$ such that $\mathbf{u} = \sum_{i=1}^n \theta_i \mathbf{x}_i$ and $\mathbf{v} = \sum_{i=1}^n \varphi_i \mathbf{x}_i$. We then have $\mathbf{u}^\mathsf{T} X \mathbf{v} = \sum_{i,j=1}^n \theta_i \varphi_i \mathbf{x}_i^\mathsf{T} X \mathbf{x}_j \geq 0$, completing the proof. □

Lemma 9.3. *Consider a partition $\mathcal{P}$ of the standard simplex with diameter equal to δ. Then we have that*

$$\mathcal{C}^n \subseteq \mathcal{O}_\mathcal{P} \subseteq \{X \in \mathcal{S}^n \mid \mathbf{v}^\mathsf{T} X \mathbf{v} \geq -2\|X\|_2 \delta \textit{ for all } \mathbf{v} \in \Delta^S\},$$
$$\mathcal{C}^n \supseteq \mathcal{I}_\mathcal{P} \supseteq \{X \in \mathcal{S}^n \mid \mathbf{v}^\mathsf{T} X \mathbf{v} \geq \tfrac{1}{2}\|X\|_2 \delta^2 \textit{ for all } \mathbf{v} \in \Delta^S\}.$$

Proof. We shall prove each inclusion relation separately:

i. It is trivial to see that $\mathcal{C}^n \subseteq \mathcal{O}_\mathcal{P}$.

ii. We consider an arbitrary $X \in \mathcal{O}_\mathcal{P}$ and an arbitrary $\mathbf{v} \in \Delta^S$. There exists a vertex $\mathbf{u}$ of $\mathcal{P}$ such that $\|\mathbf{u} - \mathbf{v}\|_2 \leq \delta$. We then have that

$$\begin{aligned} \mathbf{v}^\mathsf{T} X \mathbf{v} &= \mathbf{u}^\mathsf{T} X \mathbf{u} + (\mathbf{u} + \mathbf{v})^\mathsf{T} X (\mathbf{v} - \mathbf{u}) \\ &\geq 0 - \|\mathbf{u} + \mathbf{v}\|_2 \|X\|_2 \|\mathbf{v} - \mathbf{u}\|_2 \\ &\geq -2\|X\|_2 \delta. \end{aligned}$$

Therefore

$$\mathcal{O}_\mathcal{P} \subseteq \{X \in \mathcal{S}^n \mid \mathbf{v}^\mathsf{T} X \mathbf{v} \geq -2\|X\|_2 \delta \text{ for all } \mathbf{v} \in \Delta^S\}.$$

iii. Consider an arbitrary $X \in \mathcal{I}_\mathcal{P}$ and $\mathbf{v} \in \Delta^S$. There exists a subsimplex Δ of $\mathcal{P}$ such that $\mathbf{v} \in \Delta$. In other words, denoting the vertices of Δ by $\mathbf{u}_1, \ldots, \mathbf{u}_n$, there exists a $\boldsymbol{\theta} \in \Delta^S$ such that $\mathbf{v} = \sum_{i=1}^n \theta_i \mathbf{u}_i$. We then have that $\mathbf{v}^\mathsf{T} X \mathbf{v} = \sum_{i,j=1}^n \theta_i \theta_j \mathbf{u}_i^\mathsf{T} X \mathbf{u}_i \geq 0$. As $\mathbf{v} \in \Delta^S$ was arbitrary, this implies that $X \in \mathcal{C}^n$. Therefore, as X was also arbitrary, we get that $\mathcal{C}^n \supseteq \mathcal{I}_\mathcal{P}$.

iv. We now consider an arbitrary $X \notin \mathcal{I}_\mathcal{P}$ and show that it is also not in the set $\{X \in \mathcal{S}^n \mid \mathbf{v}^\mathsf{T} X \mathbf{v} \geq \frac{1}{2}\|X\|_2 \delta^2 \text{ for all } \mathbf{v} \in \Delta^S\}$. If $X \notin \mathcal{O}_\mathcal{P}$ then we immediately get the required result. From now on we shall assume that $X \in \mathcal{O}_\mathcal{P}$. There exists an edge $\{\mathbf{u}, \mathbf{v}\}$ of $\mathcal{P}$ such that $\mathbf{u}^\mathsf{T} X \mathbf{v} < 0$. We then have the following, noting that $\|\mathbf{u} - \mathbf{v}\|_2 \leq \delta$.

$$\begin{aligned}
\min\{\mathbf{u}^\mathsf{T} X \mathbf{u}, \mathbf{v}^\mathsf{T} X \mathbf{v}\} &\leq \tfrac{1}{2}(\mathbf{u}^\mathsf{T} X \mathbf{u} + \mathbf{v}^\mathsf{T} X \mathbf{v}) \\
&= \tfrac{1}{2}(\mathbf{u} - \mathbf{v})^\mathsf{T} X (\mathbf{u} - \mathbf{v}) + \mathbf{u}^\mathsf{T} X \mathbf{v} \\
&< \tfrac{1}{2}\|X\|_2 \|\mathbf{u} - \mathbf{v}\|_2^2 + 0 \\
&\leq \tfrac{1}{2}\|X\|_2 \delta^2.
\end{aligned}$$

□

9.1.2 Partitioning methods

In order to go from one partition to the next partition in the sequence, we consider three alternative methods, which are given in Algorithms 9.1 to 9.3. Good introductions to Algorithms 9.1 and 9.3 are provided in [HPT00]. In this chapter, we wish to have as much freedom as possible in our choices. In Algorithms 9.1 and 9.2, the choice of what value of α to pick shall always be left open, except for the restriction that $\alpha \in [\lambda, 1-\lambda]$. Similarly for the choice of $\mathbf{w}$ in Algorithm 9.3. When in Algorithms 9.1 and 9.2 an edge $\{\mathbf{u}, \mathbf{v}\}$ is picked, we say that we are *bisecting* this edge, and when in Algorithms 9.1 and 9.3 a

---

Algorithm 9.1 Classical partitioning of a single subsimplex.

Input: A parameter $\lambda \in (0, \frac{1}{2}]$, and a partition $\mathcal{P} = \{\Delta_i \mid i \in \mathcal{I}\}$.
Output: A partition $\{\Delta_j \mid j \in \mathcal{J}\}$ which is nested in the original partition $\mathcal{P}$ with $\bigcup_{j \in \mathcal{J}} \Delta_j = \bigcup_{i \in \mathcal{I}} \Delta_i$.
1: **pick** a subsimplex $\Delta_i \in \mathcal{P}$ and an edge $\{\mathbf{u}, \mathbf{v}\} \in E(\Delta_i)$.
2: **pick** a scalar $\alpha \in [\lambda, 1-\lambda]$.
3: **let** the point $\mathbf{w} = \alpha \mathbf{u} + (1-\alpha)\mathbf{v}$.
4: **let** the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_{p-2}\} = V(\Delta_i) \setminus \{\mathbf{u}, \mathbf{v}\}$.
5: **replace** Δ_i in the partition by two new simplices $\Delta_{i,1}, \Delta_{i,2}$ such that

$$\Delta_{i,1} = \operatorname{conv}\{\mathbf{u}, \mathbf{w}, \mathbf{v}_1, \ldots, \mathbf{v}_{p-2}\}, \qquad \Delta_{i,2} = \operatorname{conv}\{\mathbf{w}, \mathbf{v}, \mathbf{v}_1, \ldots, \mathbf{v}_{p-2}\}.$$

6: **output** the resultant partition.

Algorithm 9.2 Simultaneous partitioning of all subsimplices with a common edge.

Input: A parameter $\lambda \in (0, \frac{1}{2}]$, and a partition $\mathcal{P} = \{\Delta_i \mid i \in \mathcal{I}\}$.
Output: A partition $\{\Delta_j \mid j \in \mathcal{J}\}$ which is nested in the original partition $\mathcal{P}$ with $\bigcup_{j\in\mathcal{J}} \Delta_j = \bigcup_{i\in\mathcal{I}} \Delta_i$.

pick an edge $\{\mathbf{u}, \mathbf{v}\} \in E(\mathcal{P})$.
pick a scalar $\alpha \in [\lambda, 1 - \lambda]$.
let the point $\mathbf{w} = \alpha\mathbf{u} + (1 - \alpha)\mathbf{v}$.
for $i \in \mathcal{I}$ such that $\{\mathbf{u}, \mathbf{v}\} \in E(\Delta_i)$ **do**
  let the set $\{\mathbf{v}_1, \dots, \mathbf{v}_{p-2}\} = V(\Delta_i) \setminus \{\mathbf{u}, \mathbf{v}\}$.
  replace Δ_i in the partition by two new simplices $\Delta_{i,1}, \Delta_{i,2}$ such that

$$\Delta_{i,1} = \operatorname{conv}\{\mathbf{u}, \mathbf{w}, \mathbf{v}_1, \dots, \mathbf{v}_{p-2}\}, \qquad \Delta_{i,2} = \operatorname{conv}\{\mathbf{w}, \mathbf{v}, \mathbf{v}_1, \dots, \mathbf{v}_{p-2}\}.$$

end for
output the resultant partition.

subsimplex Δ_i is picked then we say that we are *partitioning* this subsimplex. As previously mentioned, it is often desirable that the sequence of partitions is exhaustive, and we shall look at which algorithms and restrictions on choices ensure this, as well as which ones do not.

Algorithm 9.1 is a commonly used method for partitioning [BD08, Hor97, Kea78]. However, from the counter-examples in the following section, we shall see that we do not get much freedom in the choice of $\{\mathbf{u}, \mathbf{v}\}$ if we wish to guarantee that the sequence of partitions is exhaustive. This is because, with this method, the same edge in two different subsimplices should really be considered as two separate edges due to the fact that they are bisected separately. For this reason we introduce Algorithm 9.2, which does not have this problem, and so, as we will see later, gives more freedom. In fact, in such cases as that in the previous subsection, this is a natural method to use, as the approximations are dependent on the edges and vertices, rather than the simplices directly.

In Algorithm 9.3 we consider radial subdivisions, also referred to in the literature as ω-subdivisions, which provides another commonly used method for partitioning [BE12, Hor76, Tuy91a, TH88]. The conditions on $\mathbf{w}$ are in fact a slight adaptation of the ρ-eccentricity condition from [Tuy91a]. In this we were restricted to $\rho \in (0, 1)$ and $\|\mathbf{w} - \mathbf{v}_j\|_2 < \rho d(\Delta_i)$ for all j. We have relaxed the restriction on ρ (although we shall later show that $\rho = 1$ should not be chosen), and added the restriction that $\mathbf{w}$ is not one of the vertices of Δ_i. In general, if ρ is too small, then it is possible that no $\mathbf{w}$ will exist satisfying this condition. However, for $\rho \geq \sqrt{3}/2$, there will exist a point obeying this condition. This is because, if we let $\mathbf{w}$ equal the midpoint of the longest edge of Δ_i, then it was shown in [Tuy91b, TH88] that $\|\mathbf{w} - \mathbf{v}\|_2 \leq \frac{1}{2}\sqrt{3}d(\Delta_i)$ for

Algorithm 9.3 Radial subdivision of a single subsimplex.

Input: A parameter $\rho \in (0, 1]$ and a partition $\mathcal{P} = \{\Delta_i \mid i \in \mathcal{I}\}$.
Output: A partition $\{\Delta_j \mid j \in \mathcal{J}\}$ which is nested in the original partition $\mathcal{P}$ with $\bigcup_{j\in\mathcal{J}} \Delta_j = \bigcup_{i\in\mathcal{I}} \Delta_i$.

pick a subsimplex $\Delta_i = \operatorname{conv}\{\mathbf{v}_1, \dots, \mathbf{v}_p\} \in \mathcal{P}$.
pick a point $\mathbf{w} \in \Delta_i \setminus V(\Delta_i)$ such that $\|\mathbf{w} - \mathbf{v}_j\|_2 < \rho d(\mathcal{P})$ for all j.
remove the simplex Δ_i from the partition.
for $j \in \{1, \dots, p\}$ **do**
  let the simplex $\Delta_{i,j} = \operatorname{conv}\{\mathbf{v}_1, \dots, \mathbf{v}_{j-1}, \mathbf{w}, \mathbf{v}_{j+1}, \dots, \mathbf{v}_p\}$.
  if $\dim(\Delta_{i,j}) = p - 1$ **then** add the simplex $\Delta_{i,j}$ to the partition.
end for
output the resultant partition.

all $\mathbf{v} \in \Delta_i$. This can also be seen using Lemma 9.5 from the following section.

Often in the literature, when analysing partitions, instead of considering the partition as a whole we consider sequences of subsimplices $\Delta_i \in \mathcal{P}_i$ for all $i \in \mathbb{Z}_+$ such that $\{\Delta_i \mid i \in \mathbb{Z}_+\}$ is a sequence of nested simplices. Conditions are then included to ensure that such a sequence is exhaustive. If all such sequences are then guaranteed to be exhaustive, then this would imply that the sequence of partitions is exhaustive. Although doing this can make the calculations for proving exhaustivity easier, in this chapter we will mainly be considering the partitions as a whole rather than sequences of subsimplices for two reasons. The first is that, in practice, keeping track of all the sequences and ensuring that the conditions hold for them can be computationally cumbersome. Secondly, Algorithm 9.2 acts in general on multiple subsimplices at once, and thus this type of analysis is insufficient.

9.2 Counter-examples

One simple choice for $\{\mathbf{u}, \mathbf{v}\}$ in Algorithms 9.1 and 9.2 is such that it is one of the longest edges in the partition. If we consider using Algorithm 9.1, then it was shown in [Kea78] that by doing this for $\lambda = \frac{1}{2}$, we do indeed get that the sequence of partitions is exhaustive. This was extended in [Hor97] to show that this is also true for any fixed parameter $\lambda \in (0, \frac{1}{2}]$, independent of the choice of α. Now, due to the similarity of Algorithms 9.1 and 9.2, we see that this result also holds for using Algorithm 9.2. This result is however of limited practical use as it does not give much freedom in the choice of $\{\mathbf{u}, \mathbf{v}\}$. Instead, in many practical applications, $\{\mathbf{u}, \mathbf{v}\}$ is picked in a way that works heuristically well for the application, which we shall refer to as a *free bisection*, then every so often a *controlling bisection* is performed which is meant to ensure the required limiting result [BD08]. Similarly, Algorithm 9.3 alone can not guarantee exhaustivity, as in general we are not bisecting edges, and so we perform Algorithm 9.3 in *free partitions* using a heuristic and every so often a controlling bisection is performed using Algorithm 9.1 or 9.2. This is formalised in Algorithm 9.4. (An alternative method to this for Algorithm 9.3, connected to a ρ-dominance condition, was given in [Tuy91a]. However this involved keeping track of sequences of subsimplices, which we wish to avoid doing.)

Algorithm 9.4 Partitioning algorithm.

Input: A simplex Δ and $q \in \mathbb{Z}_{++}$ and $\lambda \in (0, \frac{1}{2}]$. Also $\rho \in (0,1]$ if using Algorithm 9.3 in step 4.

Output: A nested sequence of partitions of Δ, denoted $\{\mathcal{P}_k \mid k \in \mathbb{Z}_+\}$.

1: **Let** $\mathcal{P}_0 = \{\Delta\}$.
2: **for** $k \in \mathbb{Z}_+$ **do**
3:   **if** $k \not\equiv q-1 \pmod q$ **then**
4:     **Let** $\mathcal{P}_{k+1}$ be a nested partition of $\mathcal{P}_k$ produced by a free bisection/partition, using Algorithm 9.1, 9.2 or 9.3.
5:   **else**
6:     **Let** $\mathcal{P}_{k+1}$ be a nested partition of $\mathcal{P}_k$ produced by a controlling bisection, using Algorithm 9.1 or 9.2.
7:   **end if**
8: **end for**

The controlling bisection is ordinarily in the form of a longest edge being bisected. It is often stated that by having this as the controlling bisection, whilst leaving the choice on the free bisection/partition unrestricted, we will have that the sequence of partitions is exhaustive (see example 9.19), where, by the free partition being unrestricted, we mean that ρ is set equal to one. However, although it seems obvious that this must be true, in the following example we shall see that this is actually not in general the case.

Example 9.4. This counter-example will be for $p = q = 3$, works for Algorithms 9.1 to 9.3, and works for every possible controlling bisection scheme, not just that of bisecting the longest edge.

We start with the triangle in Fig. 9.1a, with all edge lengths equal to one. We shall describe free bisection/partition steps such that after every controlling bisection there is at least one triangle in the partition which has $\mathbf{t}$ as one of its vertices and its opposite edge being contained in the edge $\{\mathbf{u},\mathbf{v}\}$. The diameter of the partitions is then greater than or equal to the diameter of this triangle, which is in turn greater than or equal to $\frac{1}{2}\sqrt{3}$.

Suppose that after the ith controlling bisection there is such a triangle,

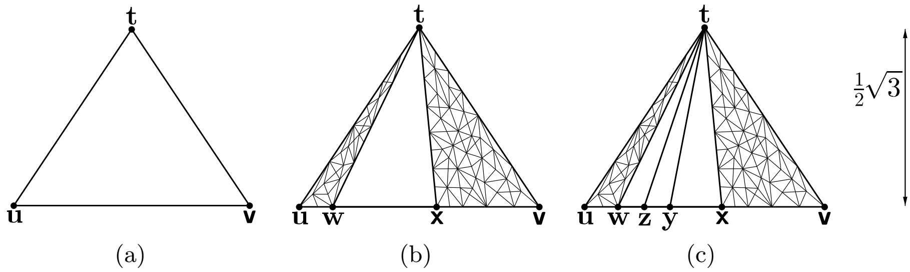


Figure 9.1: Illustration of example 9.4.

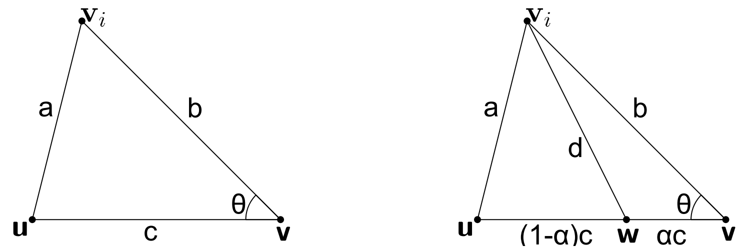
for example the triangle $\{\mathbf{t}, \mathbf{w}, \mathbf{x}\}$ in Fig. 9.1b. The original triangle can be described as the partition after the 0th controlling bisection, so this is true for $i = 0$. Now, for our two free bisection/partition steps we bisect the edge $\{\mathbf{w}, \mathbf{x}\}$ to give the vertex $\mathbf{y}$ and then the edge $\{\mathbf{w}, \mathbf{y}\}$ to give the vertex $\mathbf{z}$ (Fig. 9.1c). All three of the triangles $\{\mathbf{t}, \mathbf{w}, \mathbf{z}\}$, $\{\mathbf{t}, \mathbf{z}, \mathbf{y}\}$ and $\{\mathbf{t}, \mathbf{y}, \mathbf{x}\}$ have $\mathbf{t}$ as one of their vertices and their opposite edges being contained in the edge $\{\mathbf{u}, \mathbf{v}\}$. Now, whatever edge is bisected in the next controlling bisection, at least one of these triangles will be left untouched, and so after the $(i + 1)$th controlling bisection there will be a triangle with $\mathbf{t}$ as one of its vertices and its opposite edge being contained in the edge $\{\mathbf{u}, \mathbf{v}\}$. Thus, by induction, we have constructed a counter-example.

From this example, we see that if we wish to guarantee that the sequence of partitions is exhaustive, then we need to do something to restrict the choice of free bisections/partitions, and we shall first consider when Algorithm 9.1 or 9.2 are used in the free bisections.

The reader most likely noticed that, in example 9.4, the free bisections involved bisecting ever smaller edges. Thus one idea would be that perhaps we could pick a $\delta > 0$ and only bisect edges of length greater than or equal to δ. We would then wish for $d(\mathcal{P}_K) \leq \delta$ for some $K \in \mathbb{Z}_+$, at which point we would need to either terminate the algorithm or reduce δ. It would seem that naturally such a thing would occur even without the controlling bisections, however we shall see in example 9.6 that actually it is possible that without the controlling bisections we may not even get a reduction in $d(\mathcal{P}_k)$. Before presenting this example however, we shall first introduce the following lemma, which shall be useful in this example, along with being of use later in the chapter.

Lemma 9.5. *Consider the triangle in Fig. 9.2a, with edge lengths given by a, b, c, as labelled. (This triangle may be contained within a simplex with $p > 3$.) We partition this triangle as in Algorithm 9.2 for some $\alpha \in [0, 1]$, to produce* *the partitioned triangle in Fig. 9.2b. Then we have*

$$d^2 = \alpha a^2 + (1-\alpha)b^2 - \alpha(1-\alpha)c^2.$$



(a) Unpartitioned triangle (b) Partitioned Triangle

Figure 9.2: Triangle before and after partitioning, along with labelling.

Proof. From a standard trigonometric result, called the law of cosines, we get the following, which immediately implies the required result,

$$a^2 = b^2 + c^2 - 2bc\cos\theta \qquad \text{and} \qquad d^2 = b^2 + (\alpha c)^2 - 2b(\alpha c)\cos\theta. \qquad \square$$

We are now ready to present the example.

Example 9.6. This counter-example uses Algorithm 9.2, however it also works for using Algorithm 9.1, in which case the subsimplex $\operatorname{conv}\{\mathbf{u}, \mathbf{v}_i, \mathbf{w}\}$ should be bisected each time.

Let $p = 3$ and consider the triangle $\Delta = \operatorname{conv}\{\mathbf{u}, \mathbf{v}_0, \mathbf{w}\}$ in Fig. 9.3a, with all edge lengths equal to one. We shall describe how we can always bisect an edge of length greater than 0.65 such that the diameter of the partitions shall remain equal to one.

For all $i \in \mathbb{Z}_+$, we bisect with $\alpha = \frac{1}{2}$ to produce the new vertex $\mathbf{v}_{i+1}$, such that if i is even then the edge $\{\mathbf{u}, \mathbf{v}_i\}$ is bisected, and if i is odd then the edge $\{\mathbf{w}, \mathbf{v}_i\}$ is bisected. The first few steps of this are depicted in Fig. 9.3.

If we let l_i be the length of the edge bisected to produce vertex $\mathbf{v}_{i+1}$, then using Lemma 9.5, it can be shown that

$$l_0^2 = 1, \qquad l_1^2 = \tfrac{3}{4}, \qquad \text{and} \qquad l_i^2 = \tfrac{1}{2} - \tfrac{1}{4}l_{i-1}^2 + \tfrac{1}{8}l_{i-2}^2 \quad \text{for all } i \geq 2.$$

We can then solve this recursive relation to give the following for all $i \in \mathbb{Z}_+$:

$$l_i^2 = \tfrac{7}{9}\left(\left(-\tfrac{1}{2}\right)^i - \tfrac{1}{7}\right)^2 + \tfrac{3}{7} > \tfrac{3}{7} > (0.65)^2.$$

However, the diameter of the partition remains equal to one as neither of the edges $\{\mathbf{u}, \mathbf{w}\}$ or $\{\mathbf{v}_0, \mathbf{w}\}$ are ever bisected.

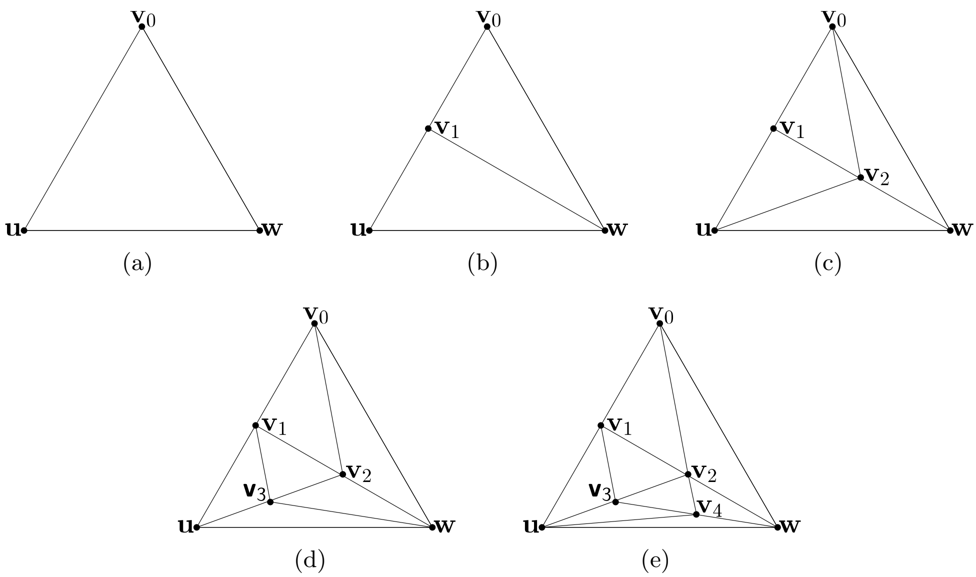


Figure 9.3: First few steps in example 9.6.

So what about combining these two approaches of only bisecting edges longer than a certain length and, every q steps, bisecting the longest edge? For this we get partial success. We shall see in the next section that if we use Algorithm 9.2 in the controlling bisections then even with a slight relaxation to this, the diameter does indeed tend towards zero. However, if instead we use Algorithm 9.1 in the controlling bisections then the following example acts as a counter-example to this approach.

Example 9.7. This counter-example is an adaptation of example 9.6.

Let $p = 4$ and consider the simplex $\Delta = \text{conv}\{\mathbf{t}, \mathbf{u}, \mathbf{v}_0, \mathbf{w}\}$, with all edge lengths equal to one. We consider $q = 2$, i.e. every other bisection is a controlling bisection. In the controlling bisections, the longest edge will be bisected using Algorithm 9.1. In the free bisections, an edge of length greater than 0.65 will be bisected using Algorithm 9.1 or 9.2. However, the diameter of the partitions shall remain equal to one throughout.

We present the following partitioning rules for $i \in \mathbb{Z}_+$, where all bisections are performed with $\alpha = \frac{1}{2}$:

- $i \equiv 0 \pmod 4$**:** Pick edge $\{\mathbf{u}, \mathbf{v}_i\}$, from subsimplex $\text{conv}\{\mathbf{t}, \mathbf{u}, \mathbf{v}_i, \mathbf{w}\}$, to bisect using Algorithm 9.1 or 9.2, producing the new vertex $\mathbf{v}_{i+2}$ and the new subsimplices $\text{conv}\{\mathbf{t}, \mathbf{u}, \mathbf{v}_{i+2}, \mathbf{w}\}$ and $\text{conv}\{\mathbf{t}, \mathbf{v}_{i+2}, \mathbf{v}_i, \mathbf{w}\}$.

- $i \equiv 1 \pmod 4$**:** Pick edge $\{\mathbf{t}, \mathbf{w}\}$, from subsimplex $\text{conv}\{\mathbf{t}, \mathbf{v}_{i+1}, \mathbf{v}_{i-1}, \mathbf{w}\}$, to bisect using Algorithm 9.1.

- $i \equiv 2 \pmod 4$**:** Pick edge $\{\mathbf{w}, \mathbf{v}_i\}$, from subsimplex $\text{conv}\{\mathbf{t}, \mathbf{u}, \mathbf{v}_i, \mathbf{w}\}$, to bisect using Algorithm 9.1 or 9.2, producing the new vertex $\mathbf{v}_{i+2}$ and the new subsimplices $\text{conv}\{\mathbf{t}, \mathbf{u}, \mathbf{v}_{i+2}, \mathbf{w}\}$ and $\text{conv}\{\mathbf{t}, \mathbf{u}, \mathbf{v}_i, \mathbf{v}_{i+2}\}$.

- $i \equiv 3 \pmod 4$**:** Pick edge $\{\mathbf{t}, \mathbf{u}\}$, from subsimplex $\text{conv}\{\mathbf{t}, \mathbf{u}, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}\}$, to bisect using Algorithm 9.1.

(Note that only some of the new subsimplices are described, however any excluded are unnecessary for understanding the example.)

For i odd, we are bisecting an edge of length equal to one, i.e. a longest edge, and for i even, if we let l_i be the length of the edge bisected, then similarly to example 9.6, we have that $l_i^2 = \frac{7}{9}\left(\left(-\frac{1}{2}\right)^{i/2} - \frac{1}{7}\right)^2 + \frac{3}{7} > (0.65)^2$.

We now return to considering Algorithm 9.3 in the free partitions. From example 9.4 we have seen that we must pick $\rho \in (0, 1)$. We next consider if simply doing this and using either Algorithm 9.1 or 9.2 in the controlling bisections to bisect the longest edge will guarantee that the sequence of partitions is exhaustive. Again we get partial success. In the next section we shall see that if we use Algorithm 9.2 in the controlling bisections then the sequence of partitions will be exhaustive. However, if instead we use Algorithm 9.1 in the controlling bisections then we can not in general guarantee exhaustivity.

Example 9.8. Similarly to the previous counter-example, we let $p = 4$, $q = 2$ and consider the simplex $\Delta = \text{conv}\{\mathbf{t}, \mathbf{u}, \mathbf{v}_0, \mathbf{w}\}$, with all edge lengths equal to one. In the controlling bisections, the longest edge will be bisected using Algorithm 9.1. In the free bisections, a subsimplex is partitioned using Algorithm 9.3 with $\rho = \frac{3}{4}$. However, the diameter of the partitions remains equal to one throughout.

We present the following partitioning rules for $i \in \mathbb{Z}_+$:

- $i \equiv 0 \pmod 2$**:** Pick the subsimplex $\text{conv}\{\mathbf{t}, \mathbf{u}, \mathbf{v}_i, \mathbf{w}\}$ to partition using Algorithm 9.3, producing the new vertex $\mathbf{v}_{i+2} = \frac{1}{4}(\mathbf{t}+\mathbf{u}+\mathbf{v}_i+\mathbf{w})$ and the new subsimplices $\text{conv}\{\mathbf{t}, \mathbf{v}_{i+2}, \mathbf{v}_i, \mathbf{w}\}$ and $\text{conv}\{\mathbf{t}, \mathbf{u}, \mathbf{v}_{i+2}, \mathbf{w}\}$, among others.

- $i \equiv 1 \pmod 2$**:** Pick the subsimplex $\text{conv}\{\mathbf{t}, \mathbf{v}_{i+1}, \mathbf{v}_{i-1}, \mathbf{w}\}$ and its edge $\{\mathbf{t}, \mathbf{w}\}$ to bisect using Algorithm 9.1.

For i odd, an edge of length equal to one is bisected, i.e. a longest edge. For i even, we are partitioning a subsimplex using Algorithm 9.3 and for $\mathbf{x} = \mathbf{t}, \mathbf{u}, \mathbf{v}_i, \mathbf{w}$ we have

$$\begin{aligned}
\|\mathbf{x} - \mathbf{v}_{i+2}\|_2 &= \|\tfrac{1}{4}(\mathbf{x} - \mathbf{t}) + \tfrac{1}{4}(\mathbf{x} - \mathbf{u}) + \tfrac{1}{4}(\mathbf{x} - \mathbf{v}_i) + \tfrac{1}{4}(\mathbf{x} - \mathbf{w})\|_2 \\
&< \tfrac{1}{4}\|\mathbf{x} - \mathbf{t}\|_2 + \tfrac{1}{4}\|\mathbf{x} - \mathbf{u}\|_2 + \tfrac{1}{4}\|\mathbf{x} - \mathbf{v}_i\|_2 + \tfrac{1}{4}\|\mathbf{x} - \mathbf{w}\|_2 \\
&\leq \tfrac{3}{4} d\left(\text{conv}\{\mathbf{t}, \mathbf{u}, \mathbf{v}_i, \mathbf{w}\}\right).
\end{aligned}$$

9.3 An Exhaustive Partitioning Scheme

In this section we will consider the following two methods of simplex partitioning.

Method 9.9. Let Δ be a simplex, $q \in \mathbb{Z}_{++}$, $\lambda \in (0, \frac{1}{2}]$ and $\eta \in (0,1]$. We consider Algorithm 9.4 for this, where

- In the free bisections, we always bisect an edge of length greater than or equal to $\eta d(\mathcal{P}_k)$ using Algorithm 9.1 or 9.2,
- In the controlling bisections, we bisect one of the longest edges using Algorithm 9.2.

Method 9.10. Let Δ be a simplex, $q \in \mathbb{Z}_{++}$, $\lambda \in (0, \frac{1}{2}]$ and $\rho \in (0,1)$. We consider Algorithm 9.4 for this, where

- In the free partitions, we partition using Algorithm 9.3,
- In the controlling bisections, we bisect one of the longest edges using Algorithm 9.2.

We shall prove the following result for these methods.

Theorem 9.11. *For Methods 9.9 and 9.10 we have that the sequence of partitions is exhaustive.*

For $q = 1$ this is already known to be true as in this case we are always bisecting the longest edge. From now on we shall consider $q \geq 2$ and shall define the following, which shall be used throughout this section and the next.

Definition 9.12. Let $\delta \in (0, d(\Delta)]$ and $q \in \mathbb{Z}$ such that $q \geq 2$. When considering Method 9.9, also define the parameter $\rho := \sqrt{1 - \lambda(1-\lambda)\eta^2}$, noting that in such case we have $\rho \in [\frac{1}{2}\sqrt{3}, 1)$. Now further define

$$p := |V(\Delta)|, \quad L := d(\Delta), \quad \gamma := \lceil \log_\rho(\delta/L) \rceil, \quad \Gamma := 2^\gamma, \quad K := \tfrac{2}{q}(\tfrac{1}{2}pq)^\Gamma.$$

In this section we shall prove the following lemma, which in turn proves Theorem 9.11. The bound in this lemma is purely there for the purpose of proving that the sequence of partitions is exhaustive, and is not tight.

Lemma 9.13. *For Methods 9.9 and 9.10, with values for the parameters given in Definition 9.12, we have $d(\mathcal{P}_k) \leq \delta$ for all $k \geq K$.*

In order to prove this lemma, we begin by considering the following lemma on partitioning a triangle.

Lemma 9.14. *Consider the triangle* $[\mathbf{u}, \mathbf{v}, \mathbf{v}_i]$*, with all edge lengths being less than or equal to* $l \in \mathbb{R}$*. (This triangle may be contained within a simplex with* $p > 3$*.) Also let* $\|\mathbf{u} - \mathbf{v}\|_2$ *be greater than or equal to* ηl *for some* $\eta \in (0, 1]$*. We partition this triangle as in Algorithm 9.1 for some* $\alpha \in [\lambda, 1 - \lambda]$*, where* $\lambda \in (0, \frac{1}{2}]$*, giving the new vertex* $\mathbf{w} = \alpha\mathbf{u} + (1 - \alpha)\mathbf{v}$ *(see Fig. 9.2b). Then all new edges produced (i.e. edges* $\{\mathbf{u}, \mathbf{w}\}$*,* $\{\mathbf{v}, \mathbf{w}\}$ *and* $\{\mathbf{v}_i, \mathbf{w}\}$*) will have lengths less than or equal to* ρl*, where* $\rho = \sqrt{1 - \lambda(1 - \lambda)\eta^2}$*.*

Proof. From Lemma 9.5 we have

$$\begin{aligned}\|\mathbf{v}_i - \mathbf{w}\|_2^2 &= \alpha\|\mathbf{u} - \mathbf{v}_i\|_2^2 + (1 - \alpha)\|\mathbf{v} - \mathbf{v}_i\|_2^2 - \alpha(1 - \alpha)\|\mathbf{u} - \mathbf{v}\|_2^2 \\ &\leq l^2 - \lambda(1 - \lambda)(\eta l)^2 = \rho^2 l^2.\end{aligned}$$

We will now complete the proof by showing that $\|\mathbf{u} - \mathbf{w}\|_2, \|\mathbf{v} - \mathbf{w}\|_2 \leq \rho l$:

$$\begin{aligned}&\|\mathbf{u} - \mathbf{w}\|_2 = (1 - \alpha)\|\mathbf{u} - \mathbf{v}\|_2 \leq (1 - \alpha)l \leq (1 - \lambda)l, \\ &\|\mathbf{v} - \mathbf{w}\|_2 = \alpha\|\mathbf{u} - \mathbf{v}\|_2 \leq \alpha l \leq (1 - \lambda)l, \\ &\quad 1 - \lambda < \sqrt{(1 - \lambda)^2 + \lambda} = \sqrt{1 - \lambda(1 - \lambda)} \leq \sqrt{1 - \lambda(1 - \lambda)\eta^2} = \rho. \qquad \square\end{aligned}$$

From this we then get the following corollary.

Corollary 9.15. *Let* $\mathcal{P}$ *be a partition and consider performing Algorithm 9.1 or 9.2 on this for* $\lambda \in (0, \frac{1}{2}]$*, with* $\|\mathbf{u} - \mathbf{v}\|_2 \geq \eta d(\mathcal{P})$*, where* $\eta \in (0, 1]$*, to produced a nested partition* $\widehat{\mathcal{P}}$*. Then any new edges produced, i.e. those in* $\widehat{\mathcal{P}}$ *but not in* $\mathcal{P}$*, have length less than or equal to* $\rho d(\mathcal{P})$*, where we let* $\rho = \sqrt{1 - \lambda(1 - \lambda)\eta^2}$*.*

If we now return to Methods 9.9 and 9.10, discussed at the beginning of this section, then we get the following, where from now on in this section we consider the values for the parameters given in Definition 9.12.

Corollary 9.16. *At any step in Methods 9.9 and 9.10, all new edges produced have length less than or equal to* $\rho d(\mathcal{P}_k)$*.*

In order to keep track of how a particular method is doing, we shall label phases in it. We say that at step k the method is in phase i if we have $\rho^{i+1}L < d(\mathcal{P}_k) \leq \rho^i L$. We then have the following lemma.

Lemma 9.17. $\widehat{p}_i := \frac{2}{q}(\frac{1}{2}pq)^{2^i}$ *is an upper bound on the number of vertices at the beginning of phase* i *(which is also a strict upper bound on the total number of steps from the beginning of the algorithm to the beginning of phase* i*).*

Proof. We have that $\widehat{p}_0 = p$, therefore the statement is true for $i = 0$. We shall now prove that if it is true for a given i, then it is also true for $i+1$, and so, by induction, we will have proven the required result.

Suppose for the sake of induction, that $\widehat{p}_i$ is an upper bound on the number of vertices at the beginning of phase i. Consider an arbitrary partition $\mathcal{P}$ in phase i. From Corollary 9.16 we get that every new edge produced when bisecting/partitioning this has length less than or equal to $\rho d(\mathcal{P}) \leq \rho^{i+1}L$. Therefore the only edges of length strictly greater than $\rho^{i+1}L$ present during this phase were there from the start of the phase, and will be the ones bisected during the controlling bisections. The total number of edges at the start of the phase is less than or equal to $\frac{1}{2}\widehat{p}_i(\widehat{p}_i - 1)$, therefore, after at most $\frac{1}{2}\widehat{p}_i(\widehat{p}_i - 1)q$ steps from the start of the phase, we have bisected all edges of length strictly greater than $\rho^{i+1}L$, and thus have left phase i. Every bisection/partition produces exactly one new vertex, therefore the total number of vertices at the beginning of phase $i+1$ is less than or equal to

$$\widehat{p}_i + \tfrac{1}{2}\widehat{p}_i(\widehat{p}_i - 1)q = \tfrac{1}{2}\widehat{p}_i^2 q + \tfrac{1}{2}\widehat{p}_i(2-q) \leq \tfrac{1}{2}\widehat{p}_i^2 q = \widehat{p}_{i+1}. \qquad \square$$

Finally we note that if $\mathcal{P}$ is the partition at the beginning of phase γ, then $d(\mathcal{P}) \leq \rho^\gamma L \leq \delta$. This then completes the proof of Lemma 9.13, which in turn proves Theorem 9.11.

If we wish to consider the dependence on δ/L more explicitly, we can note that $\gamma \leq \log_\rho(\delta\rho/L)$ and $\Gamma = 2^\gamma \leq 2^{\log_\rho(\delta\rho/L)} = (\delta\rho/L)^{\log_\rho(2)}$.

9.4 Reconsidering unrestricted free bisection

In this section we shall reconsider the use of unrestricted free bisections. We start with the following corollary of Theorem 9.11.

Corollary 9.18. *Let Δ be a simplex, $q \in \mathbb{Z}_{++}$ and $\lambda \in (0, \frac{1}{2}]$. Consider Algorithm 9.4 for this, where*

- *In the free bisections we use Algorithm 9.1 or 9.2, and we are unrestricted in our choice of edge to bisect,*
- *In the controlling bisection, we bisect one of the longest edges using Algorithm 9.2.*

Then for any $\varepsilon \in (0, d(\Delta)]$, within a finite number of steps we will have bisected an edge of length less than or equal to ε.

We shall demonstrate an application of this theorem in the following example.

Example 9.19. We return to the application of simplicial partitions to copositivity from [BD08, BD09], as described in Section 9.1.1. If we consider a matrix $A \in \mathcal{S}^n$ then

$$\begin{aligned} A \in \mathcal{I}_{\mathcal{P}} \quad &\Rightarrow \quad A \in \mathcal{C}^n, \\ A \notin \mathcal{O}_{\mathcal{P}} \quad &\Rightarrow \quad A \notin \mathcal{C}^n. \end{aligned}$$

In [BD08, BD09], the authors wished to use this to check if a matrix A is copositive, with the desire to be able to guarantee completing this check in finite time when A is in the interior of the copositive cone. For simplicity we let $\mu = \min\{\mathbf{v}^\mathsf{T} A\mathbf{v} \mid \mathbf{v} \in \Delta^S\}$. We then have that A is in the interior of the copositive cone if and only if $\mu > 0$.

The authors looked at sequences of nested partitions of the standard simplex. They defined an edge $\{\mathbf{u}, \mathbf{v}\}$ of a partition to be active if $\mathbf{u}^\mathsf{T} A\mathbf{v} < 0$, and they considered unrestricted free bisections along active edges. The paper then contains the common mistake of stating that if we apply Algorithm 9.4, using Algorithm 9.1 for all bisections, with the controlling bisections being bisecting a longest edge, then $d(\mathcal{P})$ will tend towards zero [BD08, Subsection 3.1]. However, from example 9.7, we see that this is not in general true.

Luckily this can be easily remedied. If instead we apply Algorithm 9.4, using Algorithm 9.2 for the controlling bisections, then, for $\mu > 0$, the algorithm would complete in finite time. This is because, for any active edge $\{\mathbf{u}, \mathbf{v}\}$ in a partition $\mathcal{P}$, we have

$$\tfrac{1}{2}\|A\|_2\|\mathbf{u}-\mathbf{v}\|_2^2 \geq \tfrac{1}{2}(\mathbf{u}-\mathbf{v})^\mathsf{T} A(\mathbf{u}-\mathbf{v}) = \tfrac{1}{2}(\mathbf{u}^\mathsf{T} A\mathbf{u} + \mathbf{v}^\mathsf{T} A\mathbf{v}) - \mathbf{u}^\mathsf{T} A\mathbf{v} > \mu$$

which implies that $\|\mathbf{u}-\mathbf{v}\|_2 > \sqrt{2\mu/\|A\|_2}$.

If we let $\varepsilon = \sqrt{2\mu/\|A\|_2} > 0$, then, in the free bisections, we are always bisecting active edges of length strictly greater than ε. Now, from Corollary 9.18, we see that if we have a controlling bisection of bisecting one of the longest edges using Algorithm 9.2, then it is impossible to keep bisecting edges of length greater than ε indefinitely. Therefore, within a finite number of steps we will run out of active edges to bisect, and thus have $A \in \mathcal{I}_{\mathcal{P}}$, and so the checking would be complete.

9.5 The importance of picking a fixed λ or ρ

We finish this chapter with a final example, this time to act as a reminder to the reader of the importance of picking a fixed value for λ in Algorithms 9.1 and 9.2, rather than simply limiting $\alpha \in (0, 1)$, as if this is not done then we can not guarantee exhaustivity. This example can equivalently be seen as emphasising the importance of picking $\rho < 1$ in Algorithm 9.3.

Rather than considering the partitions as a whole, we shall consider a sequence of nested simplices produced by bisections. Every other step will be a bisection at the midpoint of the longest edge, however the diameter of the simplices will not tend towards zero. Returning to the partitions as a whole, this means that even if every other step we bisected all subsimplices along one of their longest edges, then we could still not guarantee exhaustivity. This error has previously occurred in published papers, see for example [BE12, Subsection 5.2], where they used simplicial partitions to provide an alternative approximation of the copositive cone.

Example 9.20. Let $\Delta_0 = \operatorname{conv}\{\mathbf{u}_0, \mathbf{v}_0, \mathbf{w}\}$ be a triangle with all edge lengths equal to one. We shall consider triangles $\Delta_i = \operatorname{conv}\{\mathbf{u}_i, \mathbf{v}_i, \mathbf{w}\}$ such that $\{\Delta_i \mid i \in \mathbb{Z}_+\}$ is a sequence of nested triangles, defined recursively as follows:

- $i \equiv 0 \pmod 2$**:** Let $\mathbf{v}_{i+1} = \mathbf{v}_i$ and $\mathbf{u}_{i+1} = \alpha_i \mathbf{u}_i + (1-\alpha_i)\mathbf{v}_i$, where we let $\alpha_i = \frac{1}{8}\left(\frac{1}{2}\right)^{i/2}$.
- $i \equiv 1 \pmod 2$**:** Let $\mathbf{v}_{i+1} = \mathbf{u}_i$ and $\mathbf{u}_{i+1} = \frac{1}{2}(\mathbf{v}_i + \mathbf{w})$.

A diagram for two steps of this with i even is shown in Fig. 9.4.

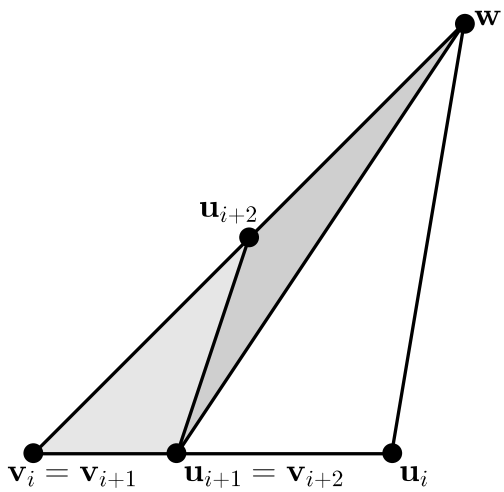


Figure 9.4: Diagram illustrating example 9.20, with i even

We show that the following hold for all $i \in \mathbb{Z}_+$:

i. $\{\mathbf{v}_i, \mathbf{w}\}$ is a longest edge of Δ_i, being the unique longest edge for $i > 0$,

ii. for i even, all edges of Δ_i have length greater than $\frac{1}{4}d(\Delta_i)$,

iii. for i even, we have $\|\mathbf{v}_i - \mathbf{w}\|_2 \geq \left(\frac{1}{4}; \frac{1}{2}\right)_{i/2} \geq \left(\frac{1}{4}; \frac{1}{2}\right)_\infty > 0.57$, where, in the literature, $(a;q)_k := \prod_{j=0}^{k-1}(1 - aq^j)$ is referred to as the *q-Pochhammer* *symbol* [Koe98].

The significance of these statements is that:

i. for i odd, we are bisecting the unique longest edge at the midpoint,

ii. for i even, we are bisecting an edge of length greater than $\frac{1}{4}d(\Delta_i)$,

iii. The diameters of the simplices are always greater than 0.57.

For $i = 0$, all edges are of length one, so the statements trivially hold. Now, for the sake of induction, assume it is true for all $i \leq 2k$, where $k \in \mathbb{Z}_+$.

By construction, $\{\mathbf{v}_{2k+1}, \mathbf{w}\}$ is the unique longest edge of Δ_{2k+1} and so the statements hold for all $i \leq 2k + 1$. We will now consider the lengths of the edges in Δ_{2k+2} in order to prove each of the statements in turn by induction:

i. We shall first prove that $\{\mathbf{v}_{2k+2}, \mathbf{w}\}$ is the unique longest edge of Δ_{2k+2}.

$$\begin{aligned}
\|\mathbf{u}_{2k+2} - \mathbf{v}_{2k+2}\|_2 &= \left\|(\tfrac{1}{2} - \alpha_{2k})(\mathbf{w} - \mathbf{v}_{2k}) + \alpha_{2k}(\mathbf{w} - \mathbf{u}_{2k})\right\|_2 \\
&< (\tfrac{1}{2} - \alpha_{2k})\|\mathbf{w} - \mathbf{v}_{2k}\|_2 + \alpha_{2k}\|\mathbf{w} - \mathbf{u}_{2k}\|_2 \\
&\leq \tfrac{1}{2}\|\mathbf{w} - \mathbf{v}_{2k}\|_2 = \|\mathbf{w} - \mathbf{u}_{2k+2}\|_2,
\end{aligned}$$

$$\begin{aligned}
\|\mathbf{w} - \mathbf{u}_{2k+2}\|_2 &= \tfrac{1}{2}\|\mathbf{w} - \mathbf{v}_{2k}\|_2 \\
&< (1 - 2\alpha_{2k})\|\mathbf{w} - \mathbf{v}_{2k}\|_2 \\
&\leq (1 - \alpha_{2k})\|\mathbf{w} - \mathbf{v}_{2k}\|_2 - \alpha_{2k}\|\mathbf{w} - \mathbf{u}_{2k}\|_2 \\
&< \|(1 - \alpha_{2k})(\mathbf{w} - \mathbf{v}_{2k}) + \alpha_{2k}(\mathbf{w} - \mathbf{u}_{2k})\|_2 \\
&= \|\mathbf{w} - \mathbf{v}_{2k+2}\|_2.
\end{aligned}$$

ii. We now prove that the length of edge $\{\mathbf{u}_{2k+2}, \mathbf{v}_{2k+2}\}$ (i.e. the shortest edge), is greater than $\frac{1}{4}$ times the length of edge $\{\mathbf{v}_{2k+2}, \mathbf{w}\}$ (i.e. the longest edge).

$$\begin{aligned}
\|\mathbf{u}_{2k+2} - \mathbf{v}_{2k+2}\|_2 &= \left\|(\tfrac{1}{2} - \alpha_{2k})(\mathbf{w} - \mathbf{v}_{2k}) + \alpha_{2k}(\mathbf{w} - \mathbf{u}_{2k})\right\|_2 \\
&> (\tfrac{1}{2} - \alpha_{2k})\|\mathbf{w} - \mathbf{v}_{2k}\|_2 - \alpha_{2k}\|\mathbf{w} - \mathbf{u}_{2k}\|_2 \\
&\geq (\tfrac{1}{2} - 2\alpha_{2k})\|\mathbf{w} - \mathbf{v}_{2k}\|_2 \\
&\geq \tfrac{1}{4}\|\mathbf{w} - \mathbf{v}_{2k}\|_2 \\
&\geq \tfrac{1}{4}\left((1 - \alpha_{2k})\|\mathbf{w} - \mathbf{v}_{2k}\|_2 + \alpha_{2k}\|\mathbf{w} - \mathbf{u}_{2k}\|_2\right) \\
&> \tfrac{1}{4}\left\|(1 - \alpha_{2k})(\mathbf{w} - \mathbf{v}_{2k}) + \alpha_{2k}(\mathbf{w} - \mathbf{u}_{2k})\right\|_2 \\
&= \tfrac{1}{4}\|\mathbf{w} - \mathbf{v}_{2k+2}\|_2.
\end{aligned}$$

iii. Finally we prove that the length of edge $\{\mathbf{v}_{2k+2}, \mathbf{w}\}$ (i.e. the longest edge) is greater than $\left(\frac{1}{4}; \frac{1}{2}\right)_{k+1}$.

$$\begin{aligned}
\|\mathbf{w} - \mathbf{v}_{2k+2}\|_2 &= \|(1-\alpha_{2k})(\mathbf{w} - \mathbf{v}_{2k}) + \alpha_{2k}(\mathbf{w} - \mathbf{u}_{2k})\|_2 \\
&> (1-\alpha_{2k})\|\mathbf{w} - \mathbf{v}_{2k}\|_2 - \alpha_{2k}\|\mathbf{w} - \mathbf{u}_{2k}\|_2 \\
&\geq (1-2\alpha_{2k})\|\mathbf{w} - \mathbf{v}_{2k}\|_2 \\
&\geq \left(1 - \frac{1}{4}\left(\frac{1}{2}\right)^k\right)\left(\frac{1}{4}; \frac{1}{2}\right)_k = \left(\frac{1}{4}; \frac{1}{2}\right)_{k+1}.
\end{aligned}$$

9.6 Conclusion

In this chapter we have seen that, for simplicial partitioning, whether a sequence of partitions is exhaustive or not can often defy our intuition. We have provided counter-examples to methods which at first glance we may believe to fulfil this requirement. Motivated by these counter-examples, a new method of partitioning was introduced which allows us to guarantee this requirement, whilst still giving us a lot of freedom.

Chapter 10

Moment Approximations*

10.1 Introduction

In this chapter, we provide outer approximation hierarchies to the set of polynomials which are nonnegative over a closed subset of the nonnegative orthant, where we may also add restrictions to the polynomials. We call such polynomials *set-semidefinite polynomials*. For example, if we restrict ourselves to homogeneous polynomials of degree 2 and consider nonnegativity over the entire nonnegative orthant, then we would, in effect, be considering the copositive cone.

We shall look at a hierarchy of outer approximations considered by Lasserre in [Las10, Las11]. Here, he considered the polynomials being nonnegative over a general closed set. We shall provide a new proof that his hierarchy provides outer approximations. This not only aids in improving our intuition on the problem, but also shows that, in the special case of nonnegativity over a closed subset of the nonnegative orthant, we can improve on his approximation.

10.1.1 Notation

We denote the ring of polynomials in the variables $\mathbf{x} = (x_1, \ldots, x_n)$ by $\mathbb{R}[\mathbf{x}]$. Elements of $\mathbb{R}[\mathbf{x}]$ are multivariate polynomials with real coefficients. Every element f of $\mathbb{R}[\mathbf{x}]$ can be written as $f(\mathbf{x}) = \sum_{\boldsymbol{\alpha}\in\mathbb{Z}^n_+} f_{\boldsymbol{\alpha}}\mathbf{x}^{\boldsymbol{\alpha}}$, where $\mathbf{x}^{\boldsymbol{\alpha}} := \prod_{i=1}^n x_i^{\alpha_i}$ (with $0^0 := 1$) and the vector $\mathbf{f} = (f_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{Z}^n_+}$ is the unique representation of f. The degree of the polynomial, denoted $\deg(f)$, is defined as the highest expo-

---

*Submitted as:

[DP13a] P.J.C. Dickinson and J. Povh. Moment approximations for set-semidefinite polynomials. *Journal of Optimization Theory and Applications*, in print. DOI: 10.1007/s10957-013-0279-7 nent appearing in f, i.e.

$$\deg(f) := \max_{\boldsymbol{\alpha} \in \mathbb{Z}_+^n} \{\boldsymbol{e}^\mathsf{T} \boldsymbol{\alpha} \mid f_{\boldsymbol{\alpha}} \neq 0\}.$$

10.1.2 Contribution

In this chapter, we consider approximating the set of polynomials which are nonnegative over a closed set $\mathcal{K} \subseteq \mathbb{R}_+^n$ (we call them *set-semidefinite polynomials*), where we may also add restrictions to the polynomials. This set is denoted as

$$\{f \in \mathcal{F} \mid f(\mathbf{x}) \geq 0 \text{ for all } \mathbf{x} \in \mathcal{K}\}, \tag{10.1}$$

where $\mathcal{F}$ represents some restrictions on f, for example restricting the degree of the polynomials or restricting them to be homogeneous. The main contributions of this chapter (and thus the paper [DP13a], whose results it is discussing) are:

- we propose an outer approximation hierarchy for the set (10.1), based on restricting moment matrices to be completely positive;
- we show that standard outer approximations of the completely positive cone give tractable outer approximation hierarchies for (10.1).

We also provide interesting new insights into the use of moments for constructing these approximations, and demonstrate the convergence of the proposed hierarchies for a couple of small scale examples.

10.2 Introduction to Moments

In this section, we recall basic definitions and concepts from moment theory. For a more detailed look at moments, and their applications to polynomial optimisation, we point the reader towards [Las10, Lau09].

Recall that a *Borel measure* on $\mathbb{R}^n$ is a nonnegative set function on Borel sets of $\mathbb{R}^n$, such that $\mu(\emptyset) = 0$ and $\mu\left(\bigcup_{i=1}^p E_i\right) = \sum_{i=1}^p \mu(E_i)$ for any countable collection of disjoint Borel sets $E_1, \ldots, E_p \subseteq \mathbb{R}^n$. A common type of Borel measure is a *probability measure*, where we recall that a Borel measure μ is a probability measure if and only if $\mu(\mathbb{R}^n) = 1$.

We define the *support* of a Borel measure μ as the minimal closed set $\mathcal{U} \subseteq \mathbb{R}^n$ such that $\mu(\mathbb{R}^n \setminus \mathcal{U}) = 0$. We denote this by $\operatorname{support}(\mu)$.

For any $\boldsymbol{\alpha} \in \mathbb{Z}_+^n$, we define the $\boldsymbol{\alpha}$ *moment* of a Borel measure μ as

$$y_{\boldsymbol{\alpha}}^\mu := \int_{\mathbb{R}^n} \mathbf{x}^{\boldsymbol{\alpha}} \mathrm{d}\mu(\mathbf{x}),$$

and we consider these moments giving the infinite dimensional vector

$$\mathbf{y}^{\mu} = (y^{\mu}_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{Z}^n_+}\,.$$

Note that the definition of moments for Borel measures is an extension of the definition of moments for probability measures. More specifically, if μ was a probability measure, then $y^{\mu}_{\boldsymbol{\alpha}}$ would be equal to the expected value of $\mathbf{x}^{\boldsymbol{\alpha}}$.

For an arbitrary Borel measure, it is not necessarily the case that all of its moments exist, and to account for this we shall let $\mathcal{M}_{\mathrm{B}}$ denote the set of Borel measures such that all their moments do in fact exist. For example, any finite Borel measure with compact support is in $\mathcal{M}_{\mathrm{B}}$, where a Borel measure μ is a *finite Borel measure* if $\mu(\mathbb{R}^n)$ is finite.

Let us now consider a polynomial $f(\mathbf{x}) = \sum_{\boldsymbol{\gamma}\in\mathbb{Z}^n_+} f_{\boldsymbol{\gamma}}\mathbf{x}^{\boldsymbol{\gamma}}$ and an infinite dimensional vector $\mathbf{y}\in\mathbb{R}^{\mathbb{Z}^n_+}$, indexed by elements in $\mathbb{Z}^n_+$. We define the *localising matrix* associated with f and $\mathbf{y}$ as

$$M(f\mathbf{y}) = \left(\sum_{\boldsymbol{\gamma}\in\mathbb{Z}^n_+} f_{\boldsymbol{\gamma}} y_{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}}\right)_{\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathbb{Z}^n_+},$$

which is an infinite order symmetric matrix, indexed by elements in $\mathbb{Z}^n_+$. Similarly, we define the localising matrix of order d by

$$M_d(f\mathbf{y}) = \left(\sum_{\boldsymbol{\gamma}\in\mathbb{Z}^n_+} f_{\boldsymbol{\gamma}} y_{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}}\right)_{\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathbb{N}^n_{\leq d}}.$$

Note that this is a symmetric matrix of finite order, indexed by elements in $\mathbb{N}^n_{\leq d}$. We also note that $M(f\mathbf{y})$ and $M_d(f\mathbf{y})$ are linearly dependent on the coefficients of f, and that $M_d(f\mathbf{y})$ is a principal submatrix of $M(f\mathbf{y})$.

10.3 Lasserre's hierarchies

In this section we consider results by Lasserre [Las11] on hierarchies of outer approximations to the set (10.1), where we are considering a general $\mathcal{K}\subseteq\mathbb{R}^n$.

A well-known theorem when considering moments in connection to polynomials is the following.

Theorem 10.1. *Consider a Borel Measure $\mu\in\mathcal{M}_{\mathrm{B}}$ and a polynomial f, such that the support of μ is equal to $\mathcal{K}\subseteq\mathbb{R}^n$ and $f(\mathbf{x})\geq 0$ for all $\mathbf{x}\in\mathcal{K}$. Letting $\mathbf{y}^{\mu}$ give the moments of μ, as described in Section 10.2, we have $M_d(f\mathbf{y}^{\mu})\in\mathcal{S}_+$ for all $d\in\mathbb{Z}_+$.*

Proof. Considering an arbitrary polynomial $g(\mathbf{x}) = \sum_{\boldsymbol{\gamma}\in\mathbb{N}^n_{\leq d}} g_{\boldsymbol{\gamma}}\mathbf{x}^{\boldsymbol{\gamma}}$, we have

$$\begin{aligned}
\mathbf{g}^{\mathsf{T}} M_d(f\mathbf{y}^{\mu})\mathbf{g} &:= \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathbb{N}^n_{\leq d}} g_{\boldsymbol{\alpha}} \left(M_d(f\mathbf{y}^{\mu})\right)_{\boldsymbol{\alpha},\boldsymbol{\beta}} g_{\boldsymbol{\beta}} \\
&= \sum_{\substack{\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathbb{N}^n_{\leq d},\\ \boldsymbol{\gamma}\in\mathbb{Z}^n_+}} g_{\boldsymbol{\alpha}} g_{\boldsymbol{\beta}} f_{\boldsymbol{\gamma}} y^{\mu}_{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}} \\
&= \sum_{\substack{\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathbb{N}^n_{\leq d},\\ \boldsymbol{\gamma}\in\mathbb{Z}^n_+}} g_{\boldsymbol{\alpha}} g_{\boldsymbol{\beta}} f_{\boldsymbol{\gamma}} \int_{\mathbb{R}^n} \mathbf{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}} \mathrm{d}\mu(\mathbf{x}) \\
&= \int_{\mathbb{R}^n} \left(\sum_{\boldsymbol{\alpha}\in\mathbb{N}^n_{\leq d}} g_{\boldsymbol{\alpha}}\mathbf{x}^{\boldsymbol{\alpha}}\right)\left(\sum_{\boldsymbol{\beta}\in\mathbb{N}^n_{\leq d}} g_{\boldsymbol{\beta}}\mathbf{x}^{\boldsymbol{\beta}}\right)\left(\sum_{\boldsymbol{\gamma}\in\mathbb{Z}^n_+} f_{\boldsymbol{\gamma}}\mathbf{x}^{\boldsymbol{\gamma}}\right) \mathrm{d}\mu(\mathbf{x}) \\
&= \int_{\mathbb{R}^n} (g(\mathbf{x}))^2 f(\mathbf{x}) \mathrm{d}\mu(\mathbf{x}) \\
&\geq 0.
\end{aligned}$$

□

Noting that all principal submatrices of a positive semidefinite matrix are also positive semidefinite, this gives us the following corollary.

Corollary 10.2. *Consider a Borel Measure $\mu \in \mathcal{M}_{\mathrm{B}}$ such that the support of μ is equal to $\mathcal{K} \subseteq \mathbb{R}^n$. Letting $\mathbf{y}^{\mu}$ give the moments of μ, as described in Section 10.2, for all $d \in \mathbb{Z}_+$ we have*

$$\begin{aligned}
\{f \in \mathcal{F} \mid f(\mathbf{x}) \geq 0 \textit{ for all } \mathbf{x} \in \mathcal{K}\} &\subseteq \left\{f \in \mathcal{F} \;\middle|\; M_{d+1}(f\mathbf{y}^{\mu}) \in \mathcal{S}_+\right\} \\
&\subseteq \left\{f \in \mathcal{F} \;\middle|\; M_d(f\mathbf{y}^{\mu}) \in \mathcal{S}_+\right\}.
\end{aligned}$$

This gives us hierarchies of outer approximations for the set that we are interested in, and we recall that $M_d(f\mathbf{y}^{\mu})$ is linearly dependent on the coefficients of f.

Naturally, we now want to know about the convergence of these hierarchies. In [Las11, Theorem 3.4], Lasserre proved that if $\mathcal{K}$ is compact and equal to the support of a Borel measure μ, then we in fact obtain

$$\{f \in \mathcal{F} \mid f(\mathbf{x}) \geq 0 \text{ for all } \mathbf{x} \in \mathcal{K}\} = \bigcap_{i\in\mathbb{Z}_+} \left\{f \in \mathcal{F} \;\middle|\; M_i(f\mathbf{y}^{\mu}) \in \mathcal{S}_+\right\},$$

and thus, in this case, the hierarchy tends towards the required set. In the case when $\mathcal{K}$ is not compact, he showed that we still get equality when there exists a finite Borel measure φ on $\mathbb{R}^n$, whose support is equal to $\mathcal{K}$, such that $\mu(B) = \int_B \exp\left(-\sum_{i=1}^n |x_i|\right) \mathrm{d}\varphi(\mathbf{x})$. Note that this is equivalent to insisting that $\int_{\mathbb{R}^n} \exp\left(\sum_{i=1}^n |x_i|\right) \mathrm{d}\mu(\mathbf{x})$ is finite, and also implies that $\mu \in \mathcal{M}_{\mathrm{B}}$ (see

Remark 10.3). Lasserre used this to show that examples of when we get equality are the multivariate normal distribution, i.e. $\mu(B) = \left(\frac{1}{2\pi}\right)^{n/2} \int_B e^{-\sum_i x_i^2/2} \mathrm{d}x$, and the multivariate exponential distribution, i.e. $\mu(B) = \int_{B\cap\mathbb{R}^n_+} e^{-\sum_i x_i} \mathrm{d}x$.

Remark 10.3. We take this opportunity to note that the proof given by Lasserre can be extended to show that we get equality whenever there exists a strictly positive vector $\boldsymbol{\lambda} \in \mathbb{R}^n$ and a finite Borel measure φ on $\mathbb{R}^n$, whose support is equal to $\mathcal{K}$, such that $\mu(B) = \int_B \exp\left(-\sum_{i=1}^n \lambda_i |x_i|\right) \mathrm{d}\varphi(\mathbf{x})$, or equivalently insisting that $\int_{\mathbb{R}^n} \exp\left(\sum_{i=1}^n \lambda_i |x_i|\right) \mathrm{d}\mu(\mathbf{x})$ is finite for some strictly positive vector $\boldsymbol{\lambda}$. This can be seen in [Las11, Subsection 3.2] by using this form for μ and using the fact that $x_i^{2k} \leq \lambda_i^{-2k}(2k)! \exp(\lambda_i|x_i|)$ for all $k \in \mathbb{Z}_+$, $\lambda_i > 0$, $x_i \in \mathbb{R}$, whilst carrying out the rest of the steps of the proof as before. Using this extension gives us a more direct proof of convergence for the multivariate exponential probability measure.

Furthermore, using the fact that for all $\lambda_i > 0$, $x_i \in \mathbb{R}$, $\alpha_i \in \mathbb{Z}_{++}$ we have that $|x_i|^{\alpha_i} \exp(-\lambda_i|x_i|) \leq \lambda_i^{-\alpha_i}\alpha_i!$, it can be observed that for such a μ as described above, and an arbitrary $\boldsymbol{\alpha} \in \mathbb{Z}^n_+$, we have

$$|y^\mu_{\boldsymbol{\alpha}}| \leq \int_{\mathbb{R}^n} \prod_{i=1}^n |x_i|^{\alpha_i} \exp(-\lambda_i|x_i|)\mathrm{d}\varphi(\mathbf{x}) \leq \varphi(\mathbb{R}^n) \prod_{i=1}^n \lambda_i^{-\alpha_i}\alpha_i!.$$

Therefore, we also have that $\mu \in \mathcal{M}_{\mathrm{B}}$, as required.

10.4 New hierarcies

In this section we consider new outer approximation hierarchies for the set that we are interested in which are based on the hierarchies in the previous section. In order to do this we first introduce some new notation.

For $n \in \mathbb{Z}_{++}$ and $d \in \mathbb{Z}_+$, we define $\mathbb{R}^{\mathbb{N}^n_{\leq d}}$ to be the set of real vectors of order $|\mathbb{N}^n_{\leq d}| = \big((n+d)!\big)/\big(n!\ d!\big)$ indexed by elements in $\mathbb{N}^n_{\leq d}$. We then define the function $\mathbf{v}_d : \mathbb{R}^n \to \mathbb{R}^{\mathbb{N}^n_{\leq d}}$ such that $\mathbf{v}_d(\mathbf{x}) := (\mathbf{x}^{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n_{\leq d}}$ for all $\mathbf{x} \in \mathbb{R}^n$.

We are now ready to present the following result, the proof of which was developed from one suggested by Prof. Dr. Monique Laurent in personal correspondence. For this we recall the notation for set-semidefinite cones from Chapter 2.

Theorem 10.4. *Consider a Borel Measure $\mu \in \mathcal{M}_{\mathrm{B}}$ and a polynomial f, such that the support of μ is equal to $\mathcal{K} \subseteq \mathbb{R}^n$ and $f(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathcal{K}$. We let $d \in \mathbb{Z}_+$ and consider a set $\mathcal{L} \subseteq \mathbb{R}^{\mathbb{N}^n_{\leq d}}$ such that $\mathbf{v}_d(\mathbf{x}) \in \mathcal{L}$ for all $\mathbf{x} \in \mathcal{K}$. Then letting $\mathbf{y}^\mu$ give the moments of μ, as described in Section 10.2, we have $M_d(f\mathbf{y}^\mu) \in \mathcal{C}^*_{\mathcal{L}}$.*

Proof. For all $X \in \mathcal{C}_{\mathcal{L}}$ we have

$$\begin{aligned}
\langle X, M_d(f\mathbf{y}^\mu)\rangle &= \sum_{\substack{\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathbb{N}^n_{\le d},\\ \boldsymbol{\gamma}\in\mathbb{Z}^n_+}} (X)_{\boldsymbol{\alpha},\boldsymbol{\beta}} f_{\boldsymbol{\gamma}} y^\mu_{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}} \\
&= \sum_{\substack{\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathbb{N}^n_{\le d},\\ \boldsymbol{\gamma}\in\mathbb{Z}^n_+}} (X)_{\boldsymbol{\alpha},\boldsymbol{\beta}} f_{\boldsymbol{\gamma}} \int_{\mathbb{R}^n} \mathbf{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}} \mathrm{d}\mu(\mathbf{x}) \\
&= \int_{\mathcal{K}} \left(\sum_{\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathbb{N}^n_{\le d}} \mathbf{x}^{\boldsymbol{\alpha}} (X)_{\boldsymbol{\alpha},\boldsymbol{\beta}} \mathbf{x}^{\boldsymbol{\beta}} \right) \left(\sum_{\boldsymbol{\gamma}\in\mathbb{Z}^n_+} f_{\boldsymbol{\gamma}} \mathbf{x}^{\boldsymbol{\gamma}} \right) \mathrm{d}\mu(\mathbf{x}) \\
&= \int_{\mathcal{K}} \underbrace{\mathbf{v}_d(\mathbf{x})^\mathsf{T} X \mathbf{v}_d(\mathbf{x})}_{\ge 0} \underbrace{f(\mathbf{x})}_{\ge 0} \mathrm{d}\mu(\mathbf{x}) \ge 0. \qquad \square
\end{aligned}$$

Noting that we have $\mathcal{C}^*_{\mathcal{L}} \subseteq \mathcal{S}_+$, this immediately gives us Theorem 10.1.

These results hold for f nonnegative over a general support of μ. In the following theorem we consider the special case of when the support of μ is contained in the nonnegative orthant.

Theorem 10.5. *Let $\mu \in \mathcal{M}_{\mathrm{B}}$, with support equal to $\mathcal{K} \subseteq \mathbb{R}^n_+$, and let $\mathbf{y}^\mu$ give its moments, as described in Section 10.2. Then for all $d \in \mathbb{Z}_+$ we have that*

$$\begin{aligned}
\{f \in \mathcal{F} \mid f(\mathbf{x}) \ge 0 \textit{ for all } \mathbf{x} \in \mathcal{K}\} &\subseteq \{f \in \mathcal{F} \mid M_{d+1}(f\mathbf{y}^\mu) \in \mathcal{C}^*\} \\
&\subseteq \{f \in \mathcal{F} \mid M_d(f\mathbf{y}^\mu) \in \mathcal{C}^*\} \\
&\subseteq \left\{f \in \mathcal{F} \;\middle|\; M_d(f\mathbf{y}^\mu) \in \mathcal{S}_+\right\}.
\end{aligned}$$

Proof. We have that $\mathcal{C}^{*m} \subseteq \mathcal{S}^m_+$ for all m, and that all principal submatrices of a completely positive matrix are also completely positive. Using these two properties it is trivial to see that

$$\begin{aligned}
\{f \in \mathcal{F} \mid M_{d+1}(f\mathbf{y}^\mu) \in \mathcal{C}^*\} &\subseteq \{f \in \mathcal{F} \mid M_d(f\mathbf{y}^\mu) \in \mathcal{C}^*\} \\
&\subseteq \left\{f \in \mathcal{F} \;\middle|\; M_d(f\mathbf{y}^\mu) \in \mathcal{S}_+\right\}.
\end{aligned}$$

Now, using Theorem 10.4 and noting that for all $\mathbf{x} \in \mathcal{K} \subseteq \mathbb{R}^n_+$ we have $\mathbf{v}_d(\mathbf{x}) \ge \mathbf{0}$, we get the following, which completes the proof.

$$\{f \in \mathcal{F} \mid f(\mathbf{x}) \ge 0 \text{ for all } \mathbf{x} \in \mathcal{K}\} \subseteq \{f \in \mathcal{F} \mid M_d(f\mathbf{y}^\mu) \in \mathcal{C}^*\}. \qquad \square$$

This gives us hierarchies of approximations which are at least as good as using the positive semidefinite cone (and thus tend towards the required set whenever the positive semidefinite approximation does). However, as we have already seen, the completely positive cone is a notoriously difficult cone to deal with. Optimising over it is in general an $\mathcal{NP}$-hard problem, and in fact even checking membership of it is an $\mathcal{NP}$-hard problem [DG11], see Chapter 3. For this reason, we would prefer to consider approximations of the completely positive cone. We recall the well-known property $\mathcal{C}^* \subseteq \mathcal{S}_+ \cap \mathcal{N} \subseteq \mathcal{S}_+$, and we shall use this relation in the following corollary in order to give a hierarchy of approximations which is a relaxation of that using the completely positive cone, but is still at least as good as that using the positive semidefinite cone.

Corollary 10.6. *Let $\mu \in \mathcal{M}_{\mathrm{B}}$, with support equal to $\mathcal{K} \subseteq \mathbb{R}^n_+$, and let $\mathbf{y}^\mu$ give its moments, as described in Section 10.2. Then for all $d \in \mathbb{Z}_+$ we have that*

$$\begin{aligned} \{f \in \mathcal{F} \mid f(\mathbf{x}) \geq 0 \text{ for all } \mathbf{x} \in \mathcal{K}\} &\subseteq \{f \in \mathcal{F} \mid M_d(f\mathbf{y}^\mu) \in \mathcal{C}^*\} \\ &\subseteq \left\{f \in \mathcal{F} \;\middle|\; M_d(f\mathbf{y}^\mu) \in \mathcal{S}_+ \cap \mathcal{N}\right\} \\ &\subseteq \left\{f \in \mathcal{F} \;\middle|\; M_d(f\mathbf{y}^\mu) \in \mathcal{S}_+\right\}, \end{aligned}$$

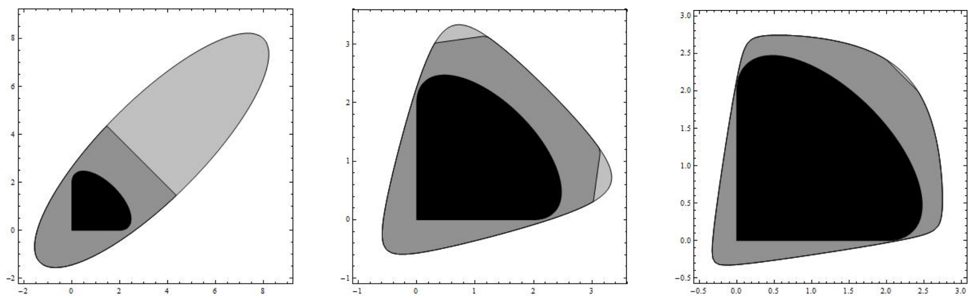
$$\left\{f \in \mathcal{F} \;\middle|\; M_{d+1}(f\mathbf{y}^\mu) \in \mathcal{S}_+ \cap \mathcal{N}\right\} \subseteq \left\{f \in \mathcal{F} \;\middle|\; M_d(f\mathbf{y}^\mu) \in \mathcal{S}_+ \cap \mathcal{N}\right\}.$$

Naturally, if the positive semidefinite hierarchy converges then all of the hierarchies will converge, where we refer the reader to our discussion on the convergence of the positive semidefinite hierarchy from Section 10.3.

10.5 Examples

We shall now use the copositive cone to consider a couple of examples of the positive semidefinite and doubly nonnegative hierarchies. For an order two symmetric matrix $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$, we define the corresponding polynomial $f_A(\mathbf{x}) = ax_1^2 + 2bx_1x_2 + cx_2^2$. We have that A is copositive if and only if $f_A(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^2_+$. Corollary 10.6 therefore provides outer approximation hierarchies for the set of copositive matrices. We shall demonstrate the effectiveness of the positive semidefinite and doubly nonnegative hierarchies by looking at a base of them, allowing us to observe the hierarchies in two dimensions. We shall also start the hierarchies with $d = 1$.

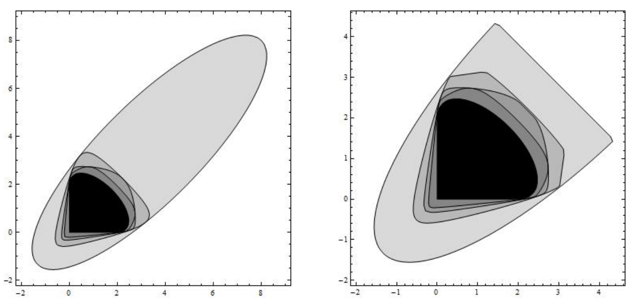
The first example that we shall look at was that demonstrated by Lasserre in [Las10]. We let μ be the multivariate exponential probability measure with rate parameter equal to the all-ones vector $\boldsymbol{e}$, i.e. $\mu(B) = \int_{B\cap\mathbb{R}^n_+} e^{-\sum_i x_i} d\mathbf{x}$, and we let $\mathbf{y}$ be the corresponding vector of moments, i.e. $y_{\boldsymbol{\alpha}} = \prod_{i=1}^n \alpha_i!$. In Fig. 10.1, we compare the hierarchies using the doubly nonnegative cone and the positive semidefinite cone for $d = 1, 2, 3$. Here, the bases are provided by intersecting with the hyperplane $\left\{ \begin{pmatrix} a & b \\ b & c \end{pmatrix} \;\middle|\; 8a + 7b + 8c = 16 \right\}$, and projecting onto the (a, c) coordinates. As we observe, although, for $d = 1$, using the doubly nonnegative cone instead of the positive semidefinite cone makes a very large difference, as d gets larger this difference is less, and in fact, for $d = 4$, we could observe no difference.



(a) d=1 (b) d=2 (c) d=3

Figure 10.1: Comparing the exponential approximations, based on the positive semidefinite cone and the doubly nonnegative cone, for the copositive cone of order two. The outer approximation (light grey) is that using the positive semidefinite cone, the next approximation (dark grey) is that using the doubly nonnegative cone, and inner most (black) is the copositive cone that they are approximating.

In Fig. 10.2 we look at how these hierarchies converge by considering $d = 1, 2, 3, 4$ and intersecting with the same hyperplane as before to get a base.



(a) Positive semidefinite (b) Doubly nonnegative

Figure 10.2: The convergence of the exponential approximations, based on the positive semidefinite cone and the doubly nonnegative cone, for the copositive cone of order two. The outer approximation (lightest grey) is for $d = 1$, with d increasing for approximations further in (darker grey). Inner most (black) is the copositive cone that they are approximating.

For our next example, we consider a compact case. We let μ be the uniform distribution over the unit box, i.e. $\mu(B) = \int_{\{\mathbf{x}\in B | \mathbf{0}\leq\mathbf{x}\leq\boldsymbol{e}\}} d\mathbf{x}$, and we let $\mathbf{y}$ be the corresponding vector of moments, i.e $y_{\boldsymbol{\alpha}} = \prod_{i=1}^{n} \frac{1}{\alpha_i+1}$. In this case, we found that for $d = 1, 2, 3, 4$, using the doubly nonnegative cone instead of the positive semidefinite cone made no difference. In Fig. 10.3, we look at how the hierarchy converges by considering $d = 1, 2, 3, 4$. Here, the base is provided by intersecting with the hyperplane $\left\{\begin{pmatrix} a & b \\ b & c \end{pmatrix} \;\middle|\; 29a + 45b + 29c = 45\right\}$ and projecting onto the (a, c) coordinates.

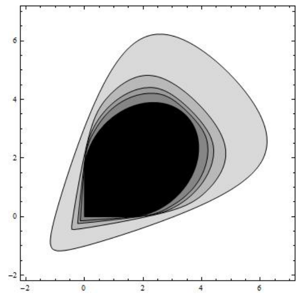


Figure 10.3: The convergence of the unit box approximation based on the positive semidefinite cone for the copositive cone of order two. The outer approximation (lightest grey) is for $d = 1$, with d increasing for approximations further in (darker grey). Inner most (black) is the copositive cone that they are approximating.

10.6 Conclusion

In this chapter, we have given new insights in to how moments can be used for approximating the set of set-semidefinite polynomials. Using this, we showed that by taking Lasserre's approximation hierarchy and replacing the positive semidefinite constraint by a completely positive constraint, we get another hierarchy which is at least as good. Due to checking membership of the completely positive cone being an $\mathcal{NP}$-hard problem, we relaxed the completely positive constraint in the approximations to the constraint of being doubly nonnegative. This again provides a hierarchy which is at least as good as that provided by the positive semidefinite cone.

In the first example, we found that using the doubly nonnegative cone instead of the positive semidefinite cone made a large difference for $d = 1$, but made less of a difference for higher d. This would suggest that for higher d, using the doubly nonnegative cone instead of the positive semidefinite cone is of little use, although it is still an open question whether this is always the case. Another open question is what the difference would be for approximations of higher order copositive cones, and whether for higher order the d makes a difference for longer.

From the second example, we observed that taking the doubly nonnegative cone instead of the positive semidefinite cone appeared to make no difference. An obvious question from this is if using the completely positive cone (or at least a better approximation of the completely positive cone) would have made any difference.

One final open question which we will finish this chapter on is whether there are any cases where we do not get convergence with the positive semidefinite cone, but we do with the completely positive cone.

Chapter 11

Sum-of-squares*

11.1 Introduction

A polynomial $f \in \mathbb{R}[\mathbf{x}]$ is defined to be *sum-of-squares* (SOS) if there exist polynomials $f_1, \ldots, f_p \in \mathbb{R}[\mathbf{x}]$ such that $f(\mathbf{x}) = \sum_{i=1}^n f_i(\mathbf{x})^2$. We then have that $f(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$.

The advantage of SOS is that we can check if a function is SOS using semidefinite optimisation. We have that a polynomial of degree less than or equal to $2d$ is SOS if and only if there exists a positive semidefinite matrix A of order $|\mathbb{N}^n_{\leq d}|$ such that $f(\mathbf{x}) = \sum_{\boldsymbol{\alpha},\boldsymbol{\beta} \in \mathbb{N}^n_{\leq d}} (A)_{\boldsymbol{\alpha},\boldsymbol{\beta}} \mathbf{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}}$.

SOS is closely tied to the idea of moments, as we shall see in the following well-known theorem. Before presenting this theorem we will first expand on the notation from localising matrices from Chapter 10.

For $d \in \mathbb{Z}_+$, $f \in \mathbb{R}[\mathbf{x}]$ and $\mathbf{y} \in \mathbb{R}^{\mathbb{Z}^n_+}$ we define

$$M_{=d}(f\mathbf{y}) := \left(\sum_{\boldsymbol{\gamma} \in \mathbb{Z}^n_+} f_{\boldsymbol{\gamma}} y_{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}} \right)_{\boldsymbol{\alpha},\boldsymbol{\beta} \in \mathbb{N}^n_{=d}} . \tag{11.1}$$

Note that this is a principal submatrix of the localising matrix of order d from Chapter 10.

For $d \in \mathbb{Z}_+$ and $\mathbf{y} \in \mathbb{R}^{\mathbb{Z}^n_+}$, we define

$$M_d(\mathbf{y}) := (y_{\boldsymbol{\alpha}+\boldsymbol{\beta}})_{\boldsymbol{\alpha},\boldsymbol{\beta} \in \mathbb{N}^n_{\leq d}}, \qquad M_{=d}(\mathbf{y}) := (y_{\boldsymbol{\alpha}+\boldsymbol{\beta}})_{\boldsymbol{\alpha},\boldsymbol{\beta} \in \mathbb{N}^n_{=d}}.$$

---

*Submitted as:

[DDGH12] P.J.C. Dickinson, M. Dür, L. Gijben and R. Hildebrand. Scaling relationship between the copositive cone and Parrilo's First level approximation. *Optimization Letters*, in print. DOI: 10.1007/s11590-012-0523-3

Theorem 11.1. *Using the notation from Chapter 10 and that introduced on the previous page, we consider the sets*

$$\begin{aligned}
\mathcal{M}_{\le d} &:= \left\{ \mathbf{y} \in \mathbb{R}^{\mathbb{N}^n_{\le 2d}} \;\middle|\; M_d(\mathbf{y}) \in \mathcal{S}_+ \right\}, \\
\mathcal{M}_{= d} &:= \left\{ \mathbf{y} \in \mathbb{R}^{\mathbb{N}^n_{= 2d}} \;\middle|\; M_{=d}(\mathbf{y}) \in \mathcal{S}_+ \right\}, \\
\Sigma_{\le d} &:= \left\{ \mathbf{f} \in \mathbb{R}^{\mathbb{N}^n_{\le 2d}} \;\middle|\; f \text{ is SOS} \right\}, \\
\Sigma_{= d} &:= \left\{ \mathbf{f} \in \mathbb{R}^{\mathbb{N}^n_{= 2d}} \;\middle|\; f \text{ is SOS} \right\}.
\end{aligned}$$

Then we have $\Sigma_{\le d} = \mathcal{M}^*_{\le d}$ *and* $\Sigma_{=d} = \mathcal{M}^*_{=d}$.

Proof. This is a well-known result, see for example [Lau09, Proposition 4.9] and [GL07, Lemma 1]. □

Furthermore, if we consider the approximation hierarchies from Chapter 10, we have the following theorem and corollary, which extends results from [Las13].

Theorem 11.2. *For* $d, s \in \mathbb{Z}_+$*, a proper cone* $\mathcal{K} \subseteq \mathcal{S}^{|\mathbb{N}^n_{\le d}|}$ *and a measure* $\mu \in \mathcal{M}_{\mathrm{B}}$*, we have*

$$\left\{ f \in \mathbb{R}^{\mathbb{N}^n_{\le s}} \;\middle|\; M_d(f\mathbf{y}^\mu) \in \mathcal{K} \right\}^* = \operatorname{cl}\left\{ \int_{\mathbb{R}^n} \mathbf{v}_s(\mathbf{x})\big(\mathbf{v}_d(\mathbf{x})^\mathsf{T} Y \mathbf{v}_d(\mathbf{x})\big)\mathrm{d}\mu(\mathbf{x}) \;\middle|\; Y \in \mathcal{K}^* \right\}.$$

Proof. Using Theorem 1.33, we get that

$$\begin{aligned}
&\left\{ f \in \mathbb{R}^{\mathbb{N}^n_{\le s}} \;\middle|\; M_d(f\mathbf{y}^\mu) \in \mathcal{K} \right\}^* \\
&\quad = \operatorname{cl}\left\{ \left(\sum_{\boldsymbol{\alpha},\boldsymbol{\beta} \in \mathbb{N}^n_{\le d}} y^\mu_{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}} Y_{\boldsymbol{\alpha}\,\boldsymbol{\beta}} \right)_{\boldsymbol{\gamma} \in \mathbb{N}^n_{\le s}} \;\middle|\; Y \in \mathcal{K}^* \right\} \\
&\quad = \operatorname{cl}\left\{ \left(\int_{\mathbb{R}^n} \sum_{\boldsymbol{\alpha},\boldsymbol{\beta} \in \mathbb{N}^n_{\le d}} \mathbf{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}} Y_{\boldsymbol{\alpha}\,\boldsymbol{\beta}} \,\mathrm{d}\mu(\mathbf{x}) \right)_{\boldsymbol{\gamma} \in \mathbb{N}^n_{\le s}} \;\middle|\; Y \in \mathcal{K}^* \right\} \\
&\quad = \operatorname{cl}\left\{ \int_{\mathbb{R}^n} \mathbf{v}_s(\mathbf{x})\big(\mathbf{v}_d(\mathbf{x})^\mathsf{T} Y \mathbf{v}_d(\mathbf{x})\big)\mathrm{d}\mu(\mathbf{x}) \;\middle|\; Y \in \mathcal{K}^* \right\}.
\end{aligned}$$ □

Corollary 11.3. *For* $s \in \mathbb{Z}_+$ *we have*

$$\left\{ f \in \mathbb{R}^{\mathbb{N}^n_{\le s}} \;\middle|\; M_d(f\mathbf{y}^\mu) \in \mathcal{S}_+ \right\}^* = \operatorname{cl}\left\{ \int_{\mathbb{R}^n} \mathbf{v}_s(\mathbf{x}) g(\mathbf{x})\mathrm{d}\mu(\mathbf{x}) \;\middle|\; \begin{array}{c} g \text{ is SOS}, \\ \deg(g) \le 2d \end{array} \right\}.$$

We thus see how, in effect, we have already used SOS to give an approximation hierarchy for the copositive cone. However, in the remainder of this chapter, we shall consider an alternative hierarchy using SOS.

11.2 Parrilo's approximation hierarchy

The Parrilo-cones were first introduced in [Par00]. For $n \in \mathbb{Z}_{++}$ and $r \in \mathbb{Z}_+$, we define the Parrilo-r cone as

$$\mathcal{K}_n^r := \left\{ A \in \mathcal{S}^n \;\middle|\; \left(\sum_{i,j=1}^n (A)_{ij} (\mathbf{x})_i^2 (\mathbf{x})_j^2 \right) \left(\sum_{i=1}^n (\mathbf{x})_i^2 \right)^r \text{ is SOS} \right\}. \tag{11.2}$$

It is trivial to see that this provides an inner approximation hierarchy for the copositive cone, and in Section 12.1 we shall see that this is in fact a convergent inner approximation hierarchy.

Parrilo showed in [Par00] that $\mathcal{K}_n^0 = \mathcal{S}_+^n + \mathcal{N}^n$, and thus from [MM62] we get that $\mathcal{C}^n = \mathcal{K}_n^0$ if and only if $n \leq 4$. Parrilo then posed the natural question "*what is the minimum n for which the $r = 1$ test is not exact?*"

In this chapter, we answer this question and show that in fact $\mathcal{C}^n \neq \mathcal{K}_n^r$ for all $r \geq 0$, $n \geq 5$.

For the order 5 case, we will also show the more surprising result that for any matrix $X \in \mathcal{C}^5$, scaling it in such a way that $(DXD)_{ii} \in \{0, 1\}$ for all i will yield $DXD \in \mathcal{K}_5^1$.

These results were in fact both conjectured by Dickinson in a reading group.

11.3 Scaling a matrix out of $\mathcal{K}_n^r$

A central ingredient in this section is the observation that given a $D = \mathrm{Diag}(\mathbf{d})$, where $\mathbf{d} \in \mathbb{R}_{++}^n$, then for any matrix class $\mathcal{X} \in \{\mathcal{C}, \mathcal{S}_+, \mathcal{N}, \mathcal{S}_+ + \mathcal{N}\}$ we have that $X \in \mathcal{X} \;\Leftrightarrow\; DXD \in \mathcal{X}$. We shall however show that property does not hold for $\mathcal{X} = \mathcal{K}_n^r$, with $r \in \mathbb{Z}_{++}$. We will show in fact that for any matrix in $X \in \mathcal{C} \setminus (\mathcal{S}_+ + \mathcal{N})$ and for any $r \in \mathbb{Z}_+$, there exists a diagonal matrix D with strictly positive diagonal such that $DXD \notin \mathcal{K}_n^r$. We shall refer to such a transformation of a matrix as a '*scaling*', as effectively we are scaling the underlying coordinate basis. We then denote the set as scalings by

$$\mathcal{D} := \{\mathrm{Diag}(\mathbf{d}) \mid \mathbf{d} \in \mathbb{R}_{++}^n\}.$$

First we shall show an auxiliary result on the relationship between the cones $\mathcal{K}_n^0$ and $\mathcal{K}_n^r$ for $r \in \mathbb{Z}_{++}$. This result was in fact proven for $r = 1$ by Dickinson and then extended for general r by Hildebrand.

Lemma 11.4. *Let $n \in \mathbb{Z}_{++}$ and $r \in \mathbb{Z}_+$. Then*

$$\{X \in \mathcal{S}^n \mid DXD \in \mathcal{K}_n^r \textit{ for all } D \in \mathcal{D}\} = \mathcal{K}_n^0.$$

Proof. Since the cone $\mathcal{K}_n^0 = \mathcal{S}_+^n + \mathcal{N}^n$ is invariant under arbitrary scalings, we have that $X \in \mathcal{K}_n^0$ implies $DXD \in \mathcal{K}_n^0 \subset \mathcal{K}_n^r$ for all $D \in \mathcal{D}$. This proves one inclusion, and the whole statement for $r = 0$.

We now prove the other inclusion for $r \in \mathbb{Z}_{++}$. Let $X \in \mathcal{S}^n$ be such that $DXD \in \mathcal{K}_n^r$ for all $D \in \mathcal{D}$. Then for all $d_1, \ldots, d_n > 0$, the polynomial

$$\left(\sum_{i,j=1}^n (X)_{ij} d_i d_j y_i^2 y_j^2\right)\left(\sum_{i=1}^n y_i^2\right)^r$$

is SOS in the variables $y_1, \ldots, y_n$. Equivalently,

$$\left(\sum_{i,j=1}^n (X)_{ij} z_i^2 z_j^2\right)\left(\sum_{i=1}^n d_i^{-1} z_i^2\right)^r$$

is SOS in the variables $z_i = \sqrt{d_i} y_i$, $i = 1, \ldots, n$. Let us now fix $d_1 = 1$ and let $d_i \to +\infty$ for $i > 1$. Since the cone of SOS is closed (this result is attributed to Robinson [Rob73]; a more accessible reference where a proof can be found is [Lau09, Section 3.8]), the limit polynomial $\left(\sum_{i,j=1}^n (X)_{ij} z_i^2 z_j^2\right)\left(z_1^2\right)^r$ is also SOS in $z_1, \ldots, z_n$, say $\left(\sum_{i,j=1}^n (X)_{ij} z_i^2 z_j^2\right) z_1^{2r} = \sum_{k=1}^N q_k^2(\mathbf{z})$. But then for all k we have that $q_k(\mathbf{z}) = 0$ whenever $z_1 = 0$. It follows that z_1 can be factored out of q_k, i.e. $\left(\sum_{i,j=1}^n (X)_{ij} z_i^2 z_j^2\right) z_1^{2(r-1)}$ is also SOS. After repeatedly carrying out this factoring out process, we arrive at the conclusion that $\left(\sum_{i,j=1}^n (X)_{ij} z_i^2 z_j^2\right)$ is SOS, i.e. $X \in \mathcal{K}_n^0$. This concludes the proof. □

Theorem 11.5. *For any $n \in \mathbb{Z}_{++}$, $r \in \mathbb{Z}_+$ and $X \in \mathcal{C}^n \setminus (\mathcal{S}_+^n + \mathcal{N}^n)$, there exists a $D \in \mathcal{D}$ such that $DXD \in \mathcal{C}^n \setminus \mathcal{K}_n^r$.*

Proof. For any $X \in \mathcal{C}^n$ and $D \in \mathcal{D}$, we have that $DXD \in \mathcal{C}^n$. Therefore we need only show that for any $X \in \mathcal{C}^n \setminus (\mathcal{S}_+^n + \mathcal{N}^n)$ there exists $D \in \mathcal{D}$ such that $DXD \notin \mathcal{K}_n^r$.

Assume for the sake of contradiction that such a D does not exist. Then for all $D \in \mathcal{D}$ we have $DXD \in \mathcal{K}_n^r$, and by Lemma 11.4 it follows that $X \in \mathcal{K}_n^0 = \mathcal{S}_+^n + \mathcal{N}^n$, a contradiction. □

Corollary 11.6. *Let $r \in \mathbb{Z}_+$ and $n \in \mathbb{Z}_{++}$. Then $\mathcal{C}^n = \mathcal{K}_n^r$ if and only if $n \leq 4$.*

11.4 Scaling a matrix into $\mathcal{K}_5^1$

In this section, we will show that in the order 5 case it is possible to scale any copositive matrix into $\mathcal{K}_5^1$. More precisely, we show that for any $X \in \mathcal{C}^5$ there exists a scaling $D \in \mathcal{D}$ such that $DXD \in \mathcal{K}_5^1$. To this end we consider matrices of the following form, which are referred to as the Hildebrand Matrices:

$$S(\boldsymbol{\theta}) := \begin{pmatrix} 1 & -\cos\theta_1 & \cos(\theta_1+\theta_2) & \cos(\theta_4+\theta_5) & -\cos\theta_5 \\ -\cos\theta_1 & 1 & -\cos\theta_2 & \cos(\theta_2+\theta_3) & \cos(\theta_5+\theta_1) \\ \cos(\theta_1+\theta_2) & -\cos\theta_2 & 1 & -\cos\theta_3 & \cos(\theta_3+\theta_4) \\ \cos(\theta_4+\theta_5) & \cos(\theta_2+\theta_3) & -\cos\theta_3 & 1 & -\cos\theta_4 \\ -\cos\theta_5 & \cos(\theta_5+\theta_1) & \cos(\theta_3+\theta_4) & -\cos\theta_4 & 1 \end{pmatrix}, \quad (11.3)$$

where $\boldsymbol{\theta} \in \Theta := \left\{\boldsymbol{\theta} \in \mathbb{R}^5_+ \;\middle|\; \boldsymbol{e}^\mathsf{T}\boldsymbol{\theta} < \pi\right\}$. These matrices were introduced by Hildebrand in the paper [Hil12] and then further investigated by Dickinson et al. in [DDGH13]. One of the main results in this latter paper was the following:

Theorem 11.7 ([DDGH13])**.** *Let $X \in \mathcal{C}^5 \setminus (\mathcal{S}^5_+ + \mathcal{N}^5)$. Then X can be decomposed as $X = R + N$, where $N \in \mathcal{N}^5$ with* $\operatorname{diag} N = \mathbf{0}$ *and $R = DP^\mathsf{T}S(\boldsymbol{\theta})PD$ for some $D \in \mathcal{D}$, a permutation matrix P and some $S(\boldsymbol{\theta})$ as defined in* (11.3) *with $\boldsymbol{\theta} \in \Theta$.*

In the paper [DDGH12], explicit certificates were found to prove the following result.

Theorem 11.8. *For all $\boldsymbol{\theta} \in \Theta$ and permutation matrices P, we have that*

$$P^\mathsf{T}S(\boldsymbol{\theta})P \in \mathcal{K}_5^1.$$

Using this we now get the following lemma.

Lemma 11.9. *Let $X \in \mathcal{S}^5$ such that* $\operatorname{diag}(X) \in \{0,1\}^5$*. Then $X \in \mathcal{C}^5$ if and only if $X \in \mathcal{K}_5^1$.*

Proof. The reverse implication trivially follows from the fact that $\mathcal{K}_5^1 \subseteq \mathcal{C}^5$.

We now consider an arbitrary $X \in \mathcal{C}^5$ such that $\operatorname{diag}(X) \in \{0,1\}^5$.

If $X \in \mathcal{S}^5_+ + \mathcal{N}^5$, then from the fact that $\mathcal{S}^5_+ + \mathcal{N}^5 = \mathcal{K}_5^0 \subseteq \mathcal{K}_5^1$, we are done.

Alternatively, if $X \in \mathcal{C}^5 \setminus (\mathcal{S}^5_+ + \mathcal{N}^5)$, then from Theorem 11.7, and noting that $\operatorname{diag}(S(\boldsymbol{\theta})) = \boldsymbol{e}$ for all $\boldsymbol{\theta}$, we see that there exists an $N \in \mathcal{N}^5$, a permutation matrix P and a $\boldsymbol{\theta} \in \Theta$ such that $X = N + P^\mathsf{T}S(\boldsymbol{\theta})P$. Now using Theorem 11.8, the fact that the Parrilo cones are convex cones and noting that $\mathcal{N}^5 \subset \mathcal{S}^5_+ + \mathcal{N}^5 \subseteq \mathcal{K}_5^1$, then implies that $X \in \mathcal{K}_5^1$ as required. □

From this, and the fact that $\mathcal{C}^5$ is invariant under scalings, we then immediately get the following two theorems.

Theorem 11.10. *Let $X \in \mathcal{S}^5$ and $D \in \mathcal{D}$ such that* $\operatorname{diag}(DXD) \in \{0,1\}^5$*. Then $X \in \mathcal{C}^5$ if and only if $DXD \in \mathcal{K}_5^1$.*

Theorem 11.11. *We have*

$$\mathcal{C}^5 = \{DXD \mid X \in \mathcal{K}_5^1,\, D \in \mathcal{D}\}.$$

11.5 The importance of scaling to binary diagonals

From Theorem 11.10, we see that if we wish to use the Parrilo-1 cone to check if an order 5 symmetric matrix is copositive, then we should scale the matrix such that all the on-diagonal entries are binary. In this section we shall look at further results suggesting the importance of scaling a matrix in this way.

Theorem 11.12 ([DDGH13]). *Let $X \in \mathcal{C}^n$ be such that $(X)_{ij} \in \{-1, +1\}$ for all $i, j = 1, \dots, n$. Then we have that $X \in \mathcal{K}_n^1$.*

Proof. We consider an arbitrary $X \in \mathcal{C}^n$ such that $(X)_{ij} \in \{-1, +1\}$ for all i, j. First note that $(X)_{ii} = 1$ for all i.

For $\mathcal{K}_n^1$, the combined proofs of Parrilo [Par00] (sufficient) and Bomze and de Klerk [BK02] (necessary) show that $X \in \mathcal{K}_n^1$ if and only if the following system of equations has a feasible solution $M^1, \dots, M^n \in \mathcal{S}^n$:

$$
\begin{aligned}
&X - M^i \in \mathcal{S}_+ && \text{for all } i = 1, \dots, n && \text{(11.4a)}\\
&(M^i)_{ii} = 0 && \text{for all } i = 1, \dots, n && \text{(11.4b)}\\
&(M^i)_{jj} + 2(M^j)_{ij} = 0 && \text{for all } i, j = 1, \dots, n \text{ s.t. } i \neq j && \text{(11.4c)}\\
&(M^i)_{jk} + (M^j)_{ik} + (M^k)_{ij} \geq 0 && \text{for all } i, j, k = 1, \dots, n \text{ s.t. } i < j < k. && \text{(11.4d)}
\end{aligned}
$$

We now consider $M^1, \dots, M^n \in \mathcal{S}^n$ given as follows,

$$M^i = X - (X\boldsymbol{e}_i)(X\boldsymbol{e}_i)^\mathsf{T} \qquad \text{for all } i = 1, \dots, n.$$

We claim that these provide a feasible solution to the system of equations (11.4a) to (11.4d), and thus a certificate for $X \in \mathcal{K}_n^1$.

From construction it is immediately apparent that for all i we have that $X - M^i$ is a positive semidefinite matrix, and so (11.4a) holds.

For all $i, j = 1, \dots, n$ we have that

$$
\begin{aligned}
(M^i)_{jj} &= (X)_{jj} - (X)_{ij}^2 = 1 - (\pm 1)^2 = 0,\\
(M^i)_{ij} &= (X)_{ij} - (X)_{ii}(X)_{ij} = (X)_{ij} - (X)_{ij} = 0.
\end{aligned}
$$

From this we immediately get that (11.4b) and (11.4c) hold.

We are now left to show that (11.4d) holds. Suppose for the sake of contradiction that there exists an $i < j < k$ such that

$$
\begin{aligned}
0 &> (M^i)_{jk} + (M^j)_{ik} + (M^k)_{ij}\\
&= (X)_{jk} + (X)_{ik} + (X)_{ij} - (X)_{ij}(X)_{ik} - (X)_{ij}(X)_{jk} - (X)_{ik}(X)_{jk}.
\end{aligned}
$$

As all the elements of X are equal to plus or minus one, it can be seen that $-1 = (X)_{jk} = (X)_{ik} = (X)_{ij}$. However, as X is copositive, we then get the following contradiction, which completes the proof:

$$0 \leq (\boldsymbol{e}_i + \boldsymbol{e}_j + \boldsymbol{e}_k)^\mathsf{T} X (\boldsymbol{e}_i + \boldsymbol{e}_j + \boldsymbol{e}_k) = -3 < 0. \qquad \square$$

Equivalent to scaling to binary would be to scale such that all the diagonal entries are either equal to zero or equal to the same positive scalar. We shall now see that further support for this type of scaling comes from the use of the Parrilo cones in approximating the stability number of a graph. As we discussed in Section 1.3, it was shown in [KP02] that the stability number $\alpha(G)$ of a graph with n nodes and adjacency matrix A_G can be computed as

$$\alpha(G) = \min\{\lambda \in \mathbb{R} \mid \lambda(I + A_G) - E \in \mathcal{C}^n\}.$$

An approximation of this can then be provided by replacing the copositive cone with one of the Parrilo cones. We now note that since A_G has a zero diagonal, we have that for any λ the diagonal entries of $\big(\lambda(I + A_G) - E\big)$ are all equal, and so the matrix is already scaled in the way that we suggest.

The support for our suggestion then comes from noting that this approximation works exceptionally well in practice. For perfect graphs we get an exact solution for all $r \geq 0$ [PVZ07, Corollary 15]. It was also conjectured in [KP02] that this approximation gives the exact solution when $r \geq \alpha(G)$. This conjecture has been proven to be true for $\alpha(G) \leq 8$ in [GL07], having previously been shown to be true for lower values of $\alpha(G)$ in [KP02, PVZ07].

11.6 Dual of the Parrilo approximations

In Theorem 11.2, we saw how the dual of the positive semidefinite moment approximation from Chapter 10 is connected to SOS. For the sake of completeness, we now briefly consider the dual of the Parrilo hierarchy.

The paper [BK02] gave the first explicit characterisation of the Parrilo cones, and we shall begin by considering some of their results.

For $\mathbf{m} \in \mathbb{Z}^n_+$ with $\boldsymbol{e}^\mathsf{T}\mathbf{m} \geq 2$ we define the following matrices

$$F_\mathbf{m} := \frac{(\boldsymbol{e}^\mathsf{T}\mathbf{m} - 2)!}{\prod_{i=1}^n (\mathbf{m})_i!} \left(\mathbf{m}\mathbf{m}^\mathsf{T} - \mathrm{Diag}(\mathbf{m})\right).$$

Using this notation, it was shown that for all $A \in \mathcal{S}^n$ and $\mathbf{z} \in \mathbb{R}^n$ we have

$$\left(\mathbf{z}^\mathsf{T} A \mathbf{z}\right)\left(\boldsymbol{e}^\mathsf{T}\mathbf{z}\right)^r = \sum_{\mathbf{m} \in \mathbb{N}^n_{=r+2}} \langle F_\mathbf{m}, A\rangle \mathbf{z}^\mathbf{m}.$$

Therefore, for all $r \in \mathbb{Z}_+$ and $n \in \mathbb{Z}_{++}$, we have

$$\mathcal{K}_n^r = \left\{ A \in \mathcal{S}^n \;\middle|\; \left(\sum_{\mathbf{m} \in \mathbb{N}^n_{=r+2}} \langle F_{\mathbf{m}}, A \rangle \mathbf{x}^{2\mathbf{m}} \right) \in \Sigma_{=r+2} \right\}.$$

We now note that

$$\left(\sum_{\mathbf{m} \in \mathbb{N}^n_{=r+2}} \langle F_{\mathbf{m}}, E \rangle \mathbf{x}^{2\mathbf{m}} \right) = \left(\sum_{\mathbf{m} \in \mathbb{N}^n_{=r+2}} \underbrace{\frac{(\boldsymbol{e}^\mathsf{T}\mathbf{m})!}{\prod_{i=1}^n (\mathbf{m})_i!}}_{>0} \mathbf{x}^{2\mathbf{m}} \right) \in \operatorname{int} \Sigma_{=r+2}.$$

Now, using Theorems 1.33 and 11.1, we find again the result of [GL07]:

$$(\mathcal{K}_n^r)^* = \left\{ \sum_{\mathbf{m} \in \mathbb{N}^n_{=r+2}} F_{\mathbf{m}} y_{2\mathbf{m}} \;\middle|\; \mathbf{y} \in \mathcal{M}_{=r+2} \right\}.$$

In order to aid in the understanding of this dual we note that

$$\begin{aligned}
\mathcal{C}^{*n} &= \operatorname{conv} \left\{ \left(\left(\textstyle\sum_{i=1}^n (\mathbf{x})_k^2 \right)^r \left((\mathbf{x})_i^2 (\mathbf{x})_j^2 \right) \right)_{i,j=1,\ldots,n} \;\middle|\; \mathbf{x} \in \mathbb{R}^n \right\} \\
&= \operatorname{conv} \left\{ \sum_{\mathbf{m} \in \mathbb{N}^n_{=r+2}} F_{\mathbf{m}} \mathbf{x}^{2\mathbf{m}} \;\middle|\; \mathbf{x} \in \mathbb{R}^n \right\} \\
&= \left\{ \sum_{\mathbf{m} \in \mathbb{N}^n_{=r+2}} F_{\mathbf{m}} y^{\mu}_{2\mathbf{m}} \;\middle|\; \mu \text{ is a Borel measure} \right\} \\
&\subseteq \left\{ \sum_{\mathbf{m} \in \mathbb{N}^n_{=r+2}} F_{\mathbf{m}} y_{2\mathbf{m}} \;\middle|\; M_{=r+2}(\mathbf{y}) \in \mathcal{S}_+ \right\}.
\end{aligned}$$

We now look at an alternative characterisation of the Parrilo cones which was presented in [PVZ07, Subsection 4.1]. We begin by letting

$$\begin{aligned}
M_{=s}(\mathbf{x}^{\mathbf{m}}\mathbf{y}) &= \left(y_{\boldsymbol{\alpha}+\boldsymbol{\beta}+\mathbf{m}} \right)_{\boldsymbol{\alpha},\boldsymbol{\beta} \in \mathbb{N}^n_{=s}}, \\
\mathcal{Y}_n^r &= \left\{ \mathbf{y} \in \mathbb{R}^{\mathbb{N}^n_{=r+2}} \;\middle|\; \begin{array}{l} \text{for all } s \in \mathbb{N}_{\leq \lfloor (r+2)/2 \rfloor},\ \mathbf{m} \in \mathbb{N}^n_{=r+2-2s} \\ \text{we have } M_{=s}(\mathbf{x}^{\mathbf{m}}\mathbf{y}) \in \mathcal{S}_+ \end{array} \right\}.
\end{aligned}$$

We have that $\mathcal{Y}_n^r$ is a closed set. Furthermore, if we let μ be the multivariate exponential probability measure, and $\mathbf{y}^\mu$ give its moments, then for all nonzero homogeneous polynomials g of degree s we have

$$\mathbf{g}^\mathsf{T} M_{=s}(\mathbf{x}^{\mathbf{m}}\mathbf{y}^\mu)\mathbf{g} = \int_{\mathbb{R}^n} \mathbf{x}^{\mathbf{m}} g(\mathbf{x})^2 \mathrm{d}\mu(\mathbf{x}) > 0,$$

and thus $M_{=s}(\mathbf{x}^{\mathbf{m}}\mathbf{y}^{\mu}) \in \operatorname{int}(\mathcal{S}_+)$.

Now using Theorems 1.33 and 11.1 we get

$$(\mathcal{Y}_n^r)^* = \left\{ \mathbf{f} \in \mathbb{R}^{\mathbb{N}^n_{=r+2}} \;\middle|\; \begin{array}{l} \text{for all } s \in \mathbb{N}_{\leq \lfloor (r+2)/2 \rfloor} \text{ and all } \mathbf{m} \in \mathbb{N}^n_{=r+2-2s} \\ \exists q_{\mathbf{m}} \in \Sigma_{=s} \text{ such that} \\ f(\mathbf{x}) = \displaystyle\sum_{s=0}^{\lfloor (r+2)/2 \rfloor} \sum_{\mathbf{m} \in \mathbb{N}^n_{=r+2-2s}} \mathbf{x}^{\mathbf{m}} q_{\mathbf{m}}(\mathbf{x}) \end{array} \right\}.$$

In [PVZ07, Subsection 4.1] it was shown that

$$\mathcal{K}_n^r = \left\{ A \in \mathcal{S}^n \;\middle|\; \exists \mathbf{f} \in (\mathcal{Y}_n^r)^* \text{ such that } \left(\mathbf{x}^\mathsf{T} A \mathbf{x}\right)\left(\boldsymbol{e}^\mathsf{T}\mathbf{x}\right)^r = f(\mathbf{x}) \right\}. \tag{11.5}$$

Similarly to before, it can be shown that for $f(\mathbf{x}) = \left(\mathbf{x}^\mathsf{T} E \mathbf{x}\right)\left(\boldsymbol{e}^\mathsf{T}\mathbf{x}\right)^r$ we have $\mathbf{f} \in \operatorname{int}(\mathcal{Y}_n^r)^*$. Therefore, using Theorems 1.33 and 11.1, we get

$$\begin{aligned} (\mathcal{K}_n^r)^* &= \left\{ \sum_{\mathbf{m} \in \mathbb{N}^n_{=r+2}} F_{\mathbf{m}} y_{\mathbf{m}} \;\middle|\; \mathbf{y} \in \mathcal{Y}_n^r \right\} \\ &= \left\{ \sum_{\mathbf{m} \in \mathbb{N}^n_{=r+2}} F_{\mathbf{m}} y_{\mathbf{m}} \;\middle|\; \begin{array}{l} \text{for all } s \in \mathbb{N}_{\leq \lfloor (r+2)/2 \rfloor},\ \mathbf{m} \in \mathbb{N}^n_{=r+2-2s} \\ \text{we have } M_{=s}(\mathbf{x}^{\mathbf{m}}\mathbf{y}) \in \mathcal{S}_+ \end{array} \right\}. \end{aligned}$$

In order to aid in the understanding of this dual we note that

$$\begin{array}{rl} \mathbf{x}^{\mathbf{m}} \geq 0 & \text{for all } \mathbf{x} \in \mathbb{R}^n_+,\ \mathbf{m} \in \mathbb{Z}^n_+, \\ M_{=s}(\mathbf{x}^{\mathbf{m}}\mathbf{y}^{\mu}) \in \mathcal{S}_+ & \text{for all } s \in \mathbb{Z}_+,\ \mathbf{m} \in \mathbb{Z}^n_+ \text{ and Borel measures } \mu \text{ such that} \\ & \text{support}(\mu) \subseteq \mathbb{R}^n_+, \end{array}$$

$$\begin{aligned} \mathcal{C}^{*n} &= \operatorname{conv}\left\{ \left(\left(\textstyle\sum_{i=1}^n (\mathbf{x})_k\right)^r ((\mathbf{x})_i (\mathbf{x})_j) \right)_{i,j=1,\dots,n} \;\middle|\; \mathbf{x} \in \mathbb{R}^n_+ \right\} \\ &= \operatorname{conv}\left\{ \sum_{\mathbf{m} \in \mathbb{N}^n_{=r+2}} F_{\mathbf{m}} \mathbf{x}^{\mathbf{m}} \;\middle|\; \mathbf{x} \in \mathbb{R}^n_+ \right\} \\ &= \left\{ \sum_{\mathbf{m} \in \mathbb{N}^n_{=r+2}} F_{\mathbf{m}} y^{\mu}_{\mathbf{m}} \;\middle|\; \begin{array}{l} \mu \text{ is a Borel measure} \\ \text{such that support}(\mu) \subseteq \mathbb{R}^n_+ \end{array} \right\} \\ &\subseteq \left\{ \sum_{\mathbf{m} \in \mathbb{N}^n_{=r+2}} F_{\mathbf{m}} y_{\mathbf{m}} \;\middle|\; \begin{array}{l} \text{for all } s \in \mathbb{N}_{\leq \lfloor (r+2)/2 \rfloor},\ \mathbf{m} \in \mathbb{N}^n_{=r+2-2s} \\ \text{we have } M_{=s}(\mathbf{x}^{\mathbf{m}}\mathbf{y}) \in \mathcal{S}_+ \end{array} \right\}. \end{aligned}$$

Chapter 12

Cones of Polynomials*

The results in this chapter shall be presented without proofs. This is primarily due to the fact that the results come from the papers [DP13b, DP13c], which were still under construction at the time of writing this thesis.

12.1 Positivstellensätze

For $n \in \mathbb{Z}_{++}$ and $r \in \mathbb{Z}_+$, we recall the formulation of the Parrilo-cones from (11.5) and consider two closely related approximation hierarchies which were introduced in [PVZ07] and [BK02] respectively:

$$\mathcal{K}_n^r = \left\{ A \in \mathcal{S}^n \;\middle|\; \begin{array}{l} \text{for all } s \in \mathbb{N}_{\le \lfloor (r+2)/2 \rfloor} \text{ and all } \mathbf{m} \in \mathbb{N}^n_{=r+2-2s}, \\ \exists \mathbf{q_m} \in \Sigma_{=s} \text{ such that} \\ \left(\mathbf{x}^\mathsf{T} A \mathbf{x}\right)\left(\boldsymbol{e}^\mathsf{T}\mathbf{x}\right)^r = \displaystyle\sum_{s=0}^{\lfloor (r+2)/2 \rfloor} \sum_{\mathbf{m} \in \mathbb{N}^n_{=r+2-2s}} \mathbf{x}^{\mathbf{m}} q_{\mathbf{m}}(\mathbf{x}) \end{array} \right\}, \tag{12.1}$$

$$\mathcal{Q}_n^r := \left\{ A \in \mathcal{S}^n \;\middle|\; \begin{array}{l} \text{for all } s \in \{0,1\} \text{ and all } \mathbf{m} \in \mathbb{N}^n_{=r+2-2s}, \\ \exists \mathbf{q_m} \in \Sigma_{=s} \text{ such that} \\ \left(\mathbf{x}^\mathsf{T} A \mathbf{x}\right)\left(\boldsymbol{e}^\mathsf{T}\mathbf{x}\right)^r = \displaystyle\sum_{s=0}^{1} \sum_{\mathbf{m} \in \mathbb{N}^n_{=r+2-2s}} \mathbf{x}^{\mathbf{m}} q_{\mathbf{m}}(\mathbf{x}) \end{array} \right\}, \tag{12.2}$$

*Submitted as:

[DP13b] P.J.C. Dickinson and J. Povh. A new convex reformulation and approximation hierarchy for polynomial optimization. Under construction.

[DP13c] P.J.C. Dickinson and J. Povh. On a generalization of Pólya's and Putinar-Vasilescu's Positivstellensätze. Preprint, submitted.

$$\mathcal{C}_n^r := \left\{ A \in \mathcal{S}^n \;\middle|\; \left(\mathbf{x}^\mathsf{T} A\mathbf{x}\right)\left(\boldsymbol{e}^\mathsf{T}\mathbf{x}\right)^r \text{ has all nonnegative coefficients} \right\} \tag{12.3}$$
$$= \left\{ A \in \mathcal{S}^n \;\middle|\; \begin{array}{l} \text{for all } \mathbf{m} \in \mathbb{N}^n_{=r+2}, \;\; \exists \mathbf{q_m} \in \Sigma_{=0} \text{ such that} \\ \left(\mathbf{x}^\mathsf{T} A\mathbf{x}\right)\left(\boldsymbol{e}^\mathsf{T}\mathbf{x}\right)^r = \displaystyle\sum_{\mathbf{m}\in\mathbb{N}^n_{=r+2}} \mathbf{x^m} q_\mathbf{m}(\mathbf{x}) \end{array} \right\}.$$

It can be seen that these are all inner approximation hierarchies for the copositive cone. Furthermore, for all r, n we have $\mathcal{C}_n^r \subseteq \mathcal{Q}_n^r \subseteq \mathcal{K}_n^r$. Therefore the results on scaling out of a cone from Section 11.3 also hold for these cones. Considering small r, we have $\mathcal{C}_n^0 = \mathcal{N}^n$ and $\mathcal{Q}_n^0 = \mathcal{K}_n^0 = \mathcal{S}_+^n + \mathcal{N}^n$ and $\mathcal{Q}_n^1 = \mathcal{K}_n^1$. Therefore the results on scaling into a cone from Section 11.4 also hold for $\mathcal{Q}_n^1$.

In order to prove that these hierarchies converge to $\mathcal{C}^n$, we use Positivstellensätze. These are basically theorems which say that if a function is strictly positive over a certain set, then there is a simple certificate which shows that it is nonnegative over this set. The following theorem is an example of a well-known positivstellensatz called Pólya's positivstellensatz:

Theorem 12.1 ([HLP88, Section 2.24])**.** *Let $f \in \mathbb{R}[\mathbf{x}]$ be a polynomial such that $f(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathbb{R}^n_+ \setminus \{\mathbf{0}\}$. Then for some $r \in \mathbb{Z}_+$, we have that all the coefficients of $(\boldsymbol{e}^\mathsf{T}\mathbf{x})^r f(\mathbf{x})$ are nonnegative.*

From this we immediately get that $\mathcal{C}_n^r$ is a convergent inner approximation hierarchy for the copositive cone, and thus so are $\mathcal{Q}_n^r$ and $\mathcal{K}_n^r$. A more direct proof on the convergence of $\mathcal{K}_n^r$ is provided by the following positivstellensatz in the case when $m = 1$ and considering the definition of $\mathcal{K}_n^r$ given in (11.2):

Theorem 12.2 ([PV99, Theorem 1])**.** *Let $m \in \mathbb{Z}_{++}$ and $f_0, \ldots, f_m \in \mathbb{R}[\mathbf{x}]$ be homogeneous polynomials of even degree on $\mathbb{R}^n$ such that $f_0(\mathbf{x}) > 0$ for all $\mathbf{x} \in \bigcap_{i=1}^m f_i^{-1}(\mathbb{R}_+) \setminus \{\mathbf{0}\}$ and $f_1(\mathbf{x}) = \mathbf{x^0}$. Then for some $r \in \mathbb{Z}_+$, there exist SOS polynomials $g_1, \ldots, g_m \in \mathbb{R}[\mathbf{x}]$ such that $(\mathbf{x}^\mathsf{T}\mathbf{x})^r f_0(\mathbf{x}) = \sum_{i=1}^m f_i(\mathbf{x}) g_i(\mathbf{x})$.*

In [DP13c], these two theorems were combined to give the following new positivstellensatz, and in this chapter we will consider results from [DP13b] on an inner approximation hierarchy based on this for generalised copositivity.

Theorem 12.3. *Let $\{f_0\} \cup \{f_i \mid i \in \mathcal{I}\} \subseteq \mathbb{R}[\mathbf{x}]$ be a set of homogeneous polynomials such that $f_0(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathbb{R}^n_+ \cap \bigcap_{i\in\mathcal{I}} f_i^{-1}(\mathbb{R}_+) \setminus \{\mathbf{0}\}$ and we have $\mathbf{x^0} \in \{f_i(\mathbf{x}) \mid i \in \mathcal{I}\}$. Then for some $r \in \mathbb{Z}_+$ there exists a subset $\mathcal{J} \subseteq \mathcal{I}$ of finite cardinality and a set of homogeneous polynomials $\{g_j \mid j \in \mathcal{J}\} \subseteq \mathbb{R}[\mathbf{x}]$, with all of their coefficients being nonnegative, such that*

$$(\boldsymbol{e}^\mathsf{T}\mathbf{x})^r f_0(\mathbf{x}) = \sum_{j\in\mathcal{J}} f_j(\mathbf{x}) g_j(\mathbf{x}).$$

12.2 Reconsidering polynomial constraints

In this section we consider how polynomial constraints can be reformulated using closed convex cones.

For $n \in \mathbb{Z}_{++}$ and $d \in \mathbb{Z}_+$, we define $\mathbb{R}^{\mathbb{N}^n_{=d}}$ to be the set of real vectors of order $|\mathbb{N}^n_{=d}| = \big((n+d-1)!\big)/\big((n-1)!\ d!\big)$ indexed by elements in $\mathbb{N}^n_{=d}$. We then define the function $\mathbf{u}_d : \mathbb{R}^n \to \mathbb{R}^{\mathbb{N}^n_{=d}}$ such that $\mathbf{u}_d(\mathbf{x}) := (\mathbf{x}^{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n_{=d}}$ for all $\mathbf{x} \in \mathbb{R}^n$.

For an arbitrary homogeneous polynomial $f \in \mathbb{R}[\mathbf{x}]$ of degree c, there exists a unique $\mathbf{f} \in \mathbb{R}^{\mathbb{N}^n_{=c}}$ such that $f(\mathbf{x}) = \langle \mathbf{f}, \mathbf{u}_c(\mathbf{x})\rangle$. This is in fact an alternative way of considering the vector forms of polynomials from Chapter 10, and similarly to before, we freely interchange between f being the polynomial and $\mathbf{f}$ being the corresponding vector in $\mathbb{R}^{\mathbb{N}^n_{=c}}$.

For $\mathbf{x} \in \mathbb{R}^n$, $i \in \mathbb{Z}_+$ and $\mathcal{F} \subseteq \mathbb{R}[\mathbf{x}]$ being a possibly infinite set of homogeneous polynomials of degree i, we have that $f(\mathbf{x}) \geq 0$ for all $f \in \mathcal{F}$ if and only if $\mathbf{u}_i(\mathbf{x}) \in \{\mathbf{f} \mid f \in \mathcal{F}\}^*$.

Using this notation, Theorem 12.3 can be written in the following form.

Theorem 12.4. *Let $\{\mathcal{Y}_i \mid i \in \mathbb{Z}_+\}$ be a set of closed convex cones such that $\mathcal{Y}_i \subseteq \mathbb{R}^{\mathbb{N}^n_{=i}}$ for all i and $\mathcal{Y}_0 = \mathbb{R}_+$. Corresponding to this we shall let $\mathcal{Y} = \{\mathbf{x} \in \mathbb{R}^n_+ \mid \mathbf{u}_i(\mathbf{x}) \in \mathcal{Y}_i$ for all $i\}$. We now consider a homogeneous polynomial $f \in \mathbb{R}[\mathbf{x}]$ of degree $d \in \mathbb{Z}_{++}$ such that $f(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathcal{Y} \setminus \{\mathbf{0}\}$. Then for some $r \in \mathbb{Z}_+$, there exists $\mathbf{g_m} \in \mathcal{Y}^*_{r+d-\boldsymbol{e}^\mathsf{T}\mathbf{m}}$ for all $\mathbf{m} \in \mathbb{N}^n_{\leq r+d}$ such that*

$$(\boldsymbol{e}^\mathsf{T}\mathbf{x})^r f(\mathbf{x}) = \sum_{\mathbf{m}\in\mathbb{N}^n_{\leq r+d}} \mathbf{x}^{\mathbf{m}} g_{\mathbf{m}}(\mathbf{x}).$$

12.3 Application to Optimisation

As in Theorem 12.4, we let $\{\mathcal{Y}_i \mid i \in \mathbb{Z}_+\}$ be a set of closed convex cones such that $\mathcal{Y}_i \subseteq \mathbb{R}^{\mathbb{N}^n_{=i}}$ for all i and $\mathcal{Y}_0 = \mathbb{R}_+$. Corresponding to this we again let $\mathcal{Y} = \{\mathbf{x} \in \mathbb{R}^n_+ \mid \mathbf{u}_i(\mathbf{x}) \in \mathcal{Y}_i$ for all $i\}$. For $d \in \mathbb{Z}_{++}$ we now define the following closed convex full-dimensional cone:

$$\mathcal{K} = \{\mathbf{f} \in \mathbb{R}^{\mathbb{N}^n_{=d}} \mid f(\mathbf{x}) \geq 0 \text{ for all } \mathbf{x} \in \mathcal{Y}\}. \tag{12.4}$$

Remark 12.5. When $d = 2$ and $\{\mathbf{u}_i(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n_+\} \subseteq \mathcal{Y}_i \subseteq \mathbb{R}^{\mathbb{N}^n_{=i}}$ for all $i \in \mathbb{Z}_+$, with $\mathcal{Y}_0 = \mathbb{R}_+$, we have that $\mathcal{K}$ is equivalent to $\mathcal{C}^n$.

It can be shown that

$$\operatorname{int}\mathcal{K} = \{\mathbf{f} \in \mathbb{R}^{\mathbb{N}^n_{=d}} \mid f(\mathbf{x}) > 0 \text{ for all } \mathbf{x} \in \mathcal{Y} \setminus \{\mathbf{0}\}\}, \tag{12.5}$$

$$\mathcal{K}^* = \operatorname{conv}\{\mathbf{u}_d(\mathbf{x}) \mid \mathbf{x} \in \mathcal{Y}\}.$$

We now consider homogeneous polynomials $g_1, g_2 \in \mathbb{R}[\mathbf{x}]$ such that *i.* $\deg(g_2) = d$,

ii. either $g_1 = 0$ or $\deg(g_1) = d$,

iii. $g_2(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathcal{Y}$,

iv. $g_1(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathcal{Y} \cap g_2^{-1}(0) \setminus \{\mathbf{0}\}$.

Using this notation, we consider the following optimisation problems:

$$\begin{aligned} \min_{\mathbf{x}} \quad & g_1(\mathbf{x}) \\ \text{s.t.} \quad & g_2(\mathbf{x}) = 1 \\ & \mathbf{u}_i(\mathbf{x}) \in \mathcal{Y}_i \qquad \text{for all } i \in \mathbb{Z}_+ \\ & \mathbf{x} \in \mathbb{R}^n_+, \end{aligned} \tag{12.6}$$

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}} \quad & \langle \mathbf{g}_1, \mathbf{y} \rangle \\ \text{s.t.} \quad & \langle \mathbf{g}_2, \mathbf{y} \rangle = 1 \\ & \mathbf{y} = \mathbf{u}_d(\mathbf{x}) \in \mathcal{K}^*, \end{aligned} \tag{12.7}$$

$$\begin{aligned} \min_{\mathbf{y}} \quad & \langle \mathbf{g}_1, \mathbf{y} \rangle \\ \text{s.t.} \quad & \langle \mathbf{g}_2, \mathbf{y} \rangle = 1 \\ & \mathbf{y} \in \mathcal{K}^*, \end{aligned} \tag{12.8}$$

$$\begin{aligned} \max_{\lambda} \quad & \lambda \\ \text{s.t.} \quad & \mathbf{g}_1 - \lambda \mathbf{g}_2 \in \mathcal{K}. \\ & \lambda \in \mathbb{R} \end{aligned} \tag{12.9}$$

A wide class of polynomial optimisation problems can be reformulated into the form given in (12.6). This includes all those with a bounded feasible set (even if there are infinitely many polynomials and no bound on the degrees of the polynomials).

Problem (12.6), is equivalent to problem (12.7). This can then be relaxed to give problem (12.8), and the dual problem to this is problem (12.9). The following theorem gives some further relations between these problems, which come from the conditions on g_1 and g_2.

Theorem 12.6 ([DP13b]). *For problems* (12.6), (12.8) *and* (12.9), *we have:*

i. $\mathrm{Val}\,(12.6) = \mathrm{Val}\,(12.8) = \mathrm{Val}\,(12.9)$.

ii. $\operatorname{Feas}(12.8) = \operatorname{conv}\{\mathbf{u}_d(\mathbf{x}) \mid \mathbf{x} \in \operatorname{Feas}(12.6)\} + \operatorname{conv}\left\{\mathbf{u}_d(\mathbf{x}) \;\middle|\; \begin{array}{l} \mathbf{x} \in \mathcal{Y}, \\ g_2(\mathbf{x}) = 0 \end{array}\right\}$.

iii. $\operatorname{Opt}(12.8) = \operatorname{conv}\{\mathbf{u}_d(\mathbf{x}) \mid \mathbf{x} \in \operatorname{Opt}(12.6)\}$.

iv. If $\operatorname{Feas}(12.6) \neq \emptyset$ *then* $\operatorname{Opt}(12.6) \neq \emptyset$. *This in turn implies that* $\operatorname{Val}(12.6)$ *is never equal to* $-\infty$.

v. For $\lambda \in \mathbb{R}$*, we have that* λ *is a strictly feasible point of* (12.9) *(i.e.* $\mathbf{g}_1 - \lambda\mathbf{g}_2 \in \operatorname{int}\mathcal{K}$*) if and only if* $\lambda < \operatorname{Val}(12.6)$.

12.4 Inner Approximation Hierarchy

In this section we shall consider a convergent inner approximation hierarchy for $\mathcal{K}$ given in (12.4), where we extend our discussion on (convergent) inner approximation hierarchies from the start of Part III for the space $\mathbb{R}^{\mathbb{N}^n_{=d}}$.

For $r \in \mathbb{Z}_+$ we define the following set:

$$\mathcal{K}_r = \left\{ \mathbf{f} \in \mathbb{R}^{\mathbb{N}^n_{=d}} \;\middle|\; \begin{array}{l} \exists \mathbf{g}_{\mathbf{m}} \in \mathcal{Y}^*_{r+d-\boldsymbol{e}^\mathsf{T}\mathbf{m}} \text{ for all } \mathbf{m} \in \mathbb{N}^n_{\le r+d} \text{ such that} \\ (\boldsymbol{e}^\mathsf{T}\mathbf{x})^r f(\mathbf{x}) = \displaystyle\sum_{\mathbf{m}\in\mathbb{N}^n_{\le r+d}} \mathbf{x}^{\mathbf{m}} g_{\mathbf{m}}(\mathbf{x}) \end{array} \right\}.$$

This clearly provides an inner approximation hierarchy for $\mathcal{K}$. Furthermore, from Theorem 12.4 and the characterisation of $\operatorname{int}\mathcal{K}$ from (12.5), we see that this is in fact a convergent inner approximation hierarchy for $\mathcal{K}$.

Remark 12.7. Reconsidering the copositive cone, as discussed in Remark 12.5, for $d = 2$, with $\mathcal{Y}_0 = \mathbb{R}_+$ and $\mathcal{Y}_i = \mathbb{R}^{\mathbb{N}^n_{=i}}$ for all $i \in \mathbb{Z}_{++}$, we get that $\mathcal{K}$ is equivalent to $\mathcal{C}^n$ and $\mathcal{K}_r$ is equivalent to $\mathcal{C}^r_n$ from (12.3).

As problem (12.9) always has strictly feasible points, it can be seen that replacing $\mathcal{K}$ with $\mathcal{K}_r$ in this problem will give a series of lower bounds on $\operatorname{Val}(12.9)$ which converge to $\operatorname{Val}(12.9)$ (see discussion on page 106).

Using a well-known result of Hilbert [Hil88], and the notation from Theorem 11.1, we have $\mathcal{M}_{=1} = \operatorname{conv}\{\mathbf{u}_2(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n\}$. Therefore, if we replace $\mathcal{Y}_2$ with $\mathcal{Y}_2 \cap \mathcal{M}_{=1}$, then this will leave $\mathcal{K}$ unchanged, whilst making the approximation $\mathcal{K}_r$ at least as good and possibly tighter.

Remark 12.8. Reconsidering the case in Remark 12.7, with $\mathcal{Y}_2$ replaced by $\mathcal{Y}_2 \cap \mathcal{M}_{=1} = \mathcal{M}_{=1}$, we get that $\mathcal{K}$ is again equivalent to $\mathcal{C}^n$ and $\mathcal{K}_r$ is equivalent to $\mathcal{Q}^r_n$ from (12.2).

The inner approximation hierarchy $\mathcal{K}_r$ has numerous theoretical advantages which are discussed in [DP13b], and the next step is to see how it works in practice.

Summary

In this thesis we studied the copositive and completely positive cones (along with some of their generalisations). These were defined respectively as follows:

$$\begin{aligned}
\mathcal{C}^n &:= \{A \in \mathcal{S}^n \mid \mathbf{x}^\mathsf{T} A\mathbf{x} \geq 0 \text{ for all } \mathbf{x} \in \mathbb{R}^n_+\},\\
\mathcal{C}^{*n} &:= \{BB^\mathsf{T} \in \mathcal{S}^n \mid B \text{ is a nonnegative matrix}\}\\
&= \operatorname{conv}\{\mathbf{b}\mathbf{b}^\mathsf{T} \mid \mathbf{b} \in \mathbb{R}^n_+\}.
\end{aligned}$$

These cones are closely related to each other by a property called duality and they are in fact mutually dual to each other.

Our motivation for studying these cones is provided by copositive optimisation. In Section 1.3 we looked at the maximum weight clique problem, which is an $\mathcal{NP}$-hard problem, and we provided a new proof for a reformulation of this into a copositive optimisation problem.

In Chapter 2 we looked at an application for a generalisation of the copositive cone in terms of a reformulation of optimisation problems with a quadratic objective function and constraints involving: linear constraints; binary constraints; set constraints. This encapsulates the well-known copositive reformulations of the standard quadratic optimisation problem and of the nonconvex quadratic binary optimisation problems as special cases.

Having considered some applications, we next wished to consider the complexity of copositivity. From the reformulation of the maximum weight clique problem into a copositive optimisation problem, we get that copositive optimisation is $\mathcal{NP}$-hard. Furthermore, in Chapter 3, we looked at how:

i. both the strong and weak membership problems for the copositive cone are $\mathcal{NP}$-hard and in the class co-$\mathcal{NP}$,

ii. both the strong and weak membership problems for the completely positive cone are $\mathcal{NP}$-hard,

iii. the weak membership problem for the completely positive cone is in the class $\mathcal{NP}$.

In spite of the complexity of checking complete positivity, in special cases this can be achieved efficiently, and in Chapter 4 we considered this for certain classes of sparse matrices. We showed that for these classes we are able to check complete positivity in linear time.

In order to improve our understanding of the copositive and completely positive cones, in Part II we considered some of their geometric properties. In Chapter 7 of this part we looked at characterisations of their interiors, whilst in Chapter 8 we looked at their faces.

Due to copositive optimisation being an $\mathcal{NP}$-hard problem, we would not expect there to be exact algorithms for solving such problems efficiently. Instead we replace the copositive cone with approximations, and this provides the subject of Part III.

The main approximations for the copositive cone can be split in to the following four categories:

i. **Simplicial partitions**
Several methods for using simplicial partitions to provide approximations to the copositive cone have previously been suggested in the literature [BD08, BE12]. In Chapter 9, we considered when we can guarantee that these methods are convergent.

ii. **Moments**
Moment theory can be applied to provide inner approximation hierarchies for the copositive cone [Las13]. In Chapter 10, we provided a new inner approximation hierarchy, based on moments, which is at least as good as those previously suggested in the literature.

iii. **Sum-of-squares**
Sum-of-squares has been used to provide an inner approximation hierarchy for the copositive cone, known as the Parrilo cones [Par00]. In Chapter 11, we looked at how scaling affects membership of the Parrilo cones. In particular we showed that for $n \geq 5$, these approximations are never exact. We also provided evidence for the importance of scaling matrices such that their on-diagonal entries are binary.

iv. **Cones of polynomials**
A final method which provides inner approximation hierarchies for the copositive cone is from using the positivstellensatz in Theorem 12.3, which was discussed in Chapter 12. This includes previous approximations of the copositive cone from [BK02, PVZ07] as special cases.

Samenvatting

In deze thesis bestudeerde we de *copositieve kegel* en de *compleet positieve kegel* (copositive cone and completely copositive cone respectively), alsmede enkele generalisaties van deze kegels. Deze kegels werden respectievelijk gedefinieerd als:

$$\begin{aligned}\mathcal{C}^n &:= \{A \in \mathcal{S}^n \mid \mathbf{x}^\mathsf{T} A\mathbf{x} \geq 0 \text{ voor iedere } \mathbf{x} \in \mathbb{R}^n_+\},\\ \mathcal{C}^{*n} &:= \{BB^\mathsf{T} \in \mathcal{S}^n \mid B \text{ een niet-negatieve matrix}\}\\ &= \operatorname{conv}\{\mathbf{b}\mathbf{b}^\mathsf{T} \mid \mathbf{b} \in \mathbb{R}^n_+\}.\end{aligned}$$

Deze kegels zijn aan elkaar verwant door middel van de eigenschap dualiteit, zij zijn dan ook elkaars duale kegel.

De motivatie voor het bestuderen van deze kegels komt van een toepassing van deze kegels, namelijk de copositieve optimalisatie. We bekeken in Sectie 1.3 het gewogen maximale kliek probleem, wat een $\mathcal{NP}$-moeilijk probleem is. We gaven een nieuw bewijs voor een herformulering van dit probleem als een copositief optimalisatie probleem.

In Hoofdstuk 2 bekeken we een toepassing van een generalisatie van de copositieve kegel, namelijk een herformulering van een klasse optimalisatie problemen met kwadratische doelfunctie en lineaire en binaire voorwaarde, alsmede voorwaarden met betrekking tot het lidmaatschap van verzamelingen. De copositieve herformulering hiervan omvat onder andere het welbekende standaard kwadratische optimalisatie probleem en het niet convexe, binaire kwadratische optimalisatie probleem als speciale gevallen.

Na het bestuderen van een aantal toepassingen wilden we de complexiteit met betrekking tot het lidmaatschap van de copositieve kegel in beschouwing nemen. Vanuit de herformulering van het gewogen maximale kliek probleem als copositief optimalisatie probleem, bleek dat copostieve optimalisatie $\mathcal{NP}$-moeilijk is. In Hoofdstuk 3 zijn we vervolgens nagegaan dat:

i. zowel het sterke als zwakke lidmaatschap probleem voor de copositieve kegel $\mathcal{NP}$-moeilijk is en tot de klasse co-$\mathcal{NP}$ behoort,

ii. zowel het sterke als zwakke lidmaatschap probleem voor de compleet positieve kegel $\mathcal{NP}$-moeilijk is,

iii. het zwakke lidmaatschap probleem voor de compleet positieve kegel tot de klasse $\mathcal{NP}$ behoort.

Ondanks de complexiteit van het vaststellen van compleet positiviteit, is het in enkele speciale gevallen toch mogelijk om dit op een efficiënte manier te bepalen. In Hoofdstuk 4 hebben we deze kwestie onderzocht voor bepaalde klassen van ijle matrices. We lieten zien dat we voor deze klassen compleet positiviteit konden vaststellen in lineaire tijd.

Om vervolgens een beter begrip van de copositieve en compleet positieve kegel te krijgen, zijn in Deel II enkele geometrische eigenschappen bekeken. In Hoofdstuk 7 van dit deel beschouwde we karakteristieken van het inwendige van beide kegels, terwijl in Hoofdstuk 8 de zijden werden besproken.

Doordat copositieve optimalisatie een $\mathcal{NP}$-moeilijk probleem is, verwachten wij niet dat er een exact algoritme bestaat dat deze problemen op een efficiënte manier oplost. In plaats daarvan laten we benaderingen de plaats innemen van de copositieve kegel, dit is het onderwerp van Deel III.

De meest belangrijke benaderingen van de copositieve kegel kunnen worden onderverdeeld in de volgende vier categorieën:

i. **Partities van het simplex**
In de literatuur zijn verschillende methode om partities van het simplex te gebruiken om benaderingen van de copositieve kegel te verkrijgen bekend [BD08, BE12]. In Hoofdstuk 9 hebben we bekeken wanneer we kunnen garanderen dat deze methoden convergeren.

ii. **Momenten**
De theorie van momenten kan worden toegepast om een hiërarchie van binnen benaderingen te vinden voor de copositieve kegel. In Hoofdstuk 10 presenteerde we een nieuwe hiërarchie van binnen benaderingen, gebaseerd op momenten, die tenminste even goed is als de hiërarchieën reeds bekend uit de literatuur.

iii. **Som van kwadraten**
De som van kwadraten is gebruikt om een hiërarchie van binnen benaderingen van de copositieve kegel te verkrijgen, deze benaderingen staan bekend als de Parrilo kegels [Par00]. In Hoofdstuk 11 bekeken we hoe schaling het lidmaatschap van de Parrilo kegels beïnvloedt. In het bijzonder lieten we zien dat voor $n \geq 5$, deze benaderingen nooit exact zijn. We verschaffen daarnaast ook nog aanwijzingen waarom het belangrijk zou kunnen zijn om matrices zo te schalen dat de elementen op de hoofddiagonaal binair zijn.

iv. **Kegels van polynomen**
Een laatste methode om hiërarchieën van binnen benaderingen voor de copositieve kegel te construeren komt van de positivstellensatz in Stelling 12.3, welke besproken werd in Hoofdstuk 12. Deze hiërarchieën omvatten zowel eerdere benaderingen van de copositieve kegel uit [BK02, PVZ07] als speciale gevallen.

Bibliography

[ABD12] Kurt M. Anstreicher, Samuel Burer, and Peter J.C. Dickinson. An algorithm for computing the cp-factorization of a completely positive matrix. *In construction*, 2012.

[AT02] Alfred Auslender and Marc Teboulle. *Asymptotic Cones and Functions in Optimization and Variational Inequalities.* Springer, 2002.

[Bau66] Leonard D. Baumert. Extreme copositive quadratic forms. *Pacific Journal of Mathematics*, 19:197–204, 1966.

[BB03] Francesco Barioli and Abraham Berman. The maximal cp-rank of rank k completely positive matrices. *Linear Algebra and its Applications*, 363:17–33, Apr 2003.

[BD08] Stefan Bundfuss and Mirjam Dür. Algorithmic copositivity detection by simplicial partition. *Linear Algebra and its Applications*, 428(7):1511–1523, Apr 2008.

[BD09] Stefan Bundfuss and Mirjam Dür. An adaptive linear approximation algorithm for copositive programs. *SIAM Journal on Optimization*, 20(1):30–53, 2009.

[BDK$^+$00] Immanuel M. Bomze, Mirjam Dür, Etienne de Klerk, Cornelis Roos, Arie J. Quist, and Tamàs Terlaky. On copositive programming and standard quadratic optimization problems. *Journal of Global Optimization*, 18:301–320, 2000.

[BE12] Immanuel Bomze and Gabriele Eichfelder. Copositivity detection by difference-of-convex decomposition and ω-subdivision. *Mathematical Programming, in print. DOI: 10.1007/s10107-012-0543-x*, 2012.

[Ber73] Abraham Berman. *Cones, Matrices and Mathematical Programming*, volume 79 of *Lecture Notes in Economics and Mathematical Systems.* Springer Verlag, 1973.

[Ber09] Dennis S. Bernstein. *Matrix Mathematics: Theory, Facts, and Formulas.* Princeton University Press, 2nd edition, 2009.

[BG88] Abraham Berman and Robert Grone. Bipartite completely positive matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 103:269–276, 1988.

[BH87] Abraham Berman and Daniel Hershkowitz. Combinatorial results on completely positive matrices. *Linear Algebra and its Applications*, 95:111–125, 1987.

[BJ10] Immanuel Bomze and Florian Jarre. A note on Burer's copositive representation of mixed-binary QPs. *Optimization Letters*, 4(3):465–472, 2010.

[BK02] Immanuel M. Bomze and Etienne de Klerk. Solving standard quadratic optimization problems via linear, semidefinite and copositive programming. *Journal of Global Optimization*, 24(2):163–185, 2002.

[Bom98] Immanuel M. Bomze. On standard quadratic optimization problems. *Journal of Global Optimization*, 13(4):369–387, 1998.

[Bom00] Immanuel M. Bomze. Linear-time copositivity detection for tridiagonal matrices and extension to block-tridiagonality. *SIAM Journal on Matrix Analysis and Applications*, 21(3):840–848, 2000.

[Bom12] Immanuel M. Bomze. Copositive optimization - recent developments and applications. *European Journal of Operational Research*, 216(3):509–520, 2012.

[BR06] Abraham Berman and Uriel G. Rothblum. A note on the computation of the CP-rank. *Linear Algebra and its Applications*, 419:1–7, 2006.

[Brø83] Arne Brøndsted. *An Introduction to Convex Polytopes.* Springer-Verlag, 1983.

[BSM03] Abraham Berman and Naomi Shaked-Monderer. *Completely Positive Matrices.* World Scientific, 2003.

[Bun09] Stefan Bundfuss. *Copositive Matrices, Copositive Programming, and Applications.* PhD thesis, TU Darmstadt, 2009.

[Bur09] Samuel Burer. On the copositive representation of binary and continuous nonconvex quadratic programs. *Mathematical Programming*, 120(2):479–495, 2009.

[Bur12] Samuel Burer. *Copositive Programming,* In: Handbook on Semidefinite, Conic and Polynomial Optimization, Miguel F. Anjos and Jean-Bernard Lasserre (Eds.), volume 166 of *International Series in Operations Research & Management Science*, pages 201–218. Springer US, 2012.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[DA13] Hongbo Dong and Kurt Anstreicher. Separating doubly nonnegative and completely positive matrices. *Mathematical Programming*, 137:131–153, 2013.

[Dan98] George Dantzig. *Linear Programming and Extensions.* Princeton University Press, Aug 1998.

[DD12] Peter J.C. Dickinson and Mirjam Dür. Linear-time complete positivity detection and decomposition of sparse matrices. *SIAM Journal on Matrix Analysis and Applications*, 33(3):701–720, 2012.

[DDGH12] Peter J.C. Dickinson, Mirjam Dür, Luuk Gijben, and Roland Hildebrand. Scaling relationship between the copositive cone and parrilo's first level approximation. *Optimization Letters, in print. DOI: 10.1007/s11590-012-0523-3*, 2012.

[DDGH13] Peter J.C. Dickinson, Mirjam Dür, Luuk Gijben, and Roland Hildebrand. Irreducible elements of the copositive cone. *Linear Algebra and its Applications, accepted. Available at http://www.optimization-online.org/DB_HTML/2012/03/3383.html*, 2013.

[DEP12] Peter J.C. Dickinson, Gabriele Eichfelder, and Janez Povh. Erratum to "On the set-semidefinite representation of nonconvex quadratic programs over arbitrary feasible sets" [Optim. Letters, 2012]. *Optimization Letters, accepted. Available at http://www.optimization-online.org/DB_HTML/2012/09/3598.html*, 2012.

[DG11] Peter J.C. Dickinson and Luuk Gijben. On the computational complexity of membership problems for the completely positive cone and its dual. *Computational Optimization and Applications, conditionally accepted. Available at http://www.optimization-online.org/DB_HTML/2011/05/3041.html*, 2011.

[Dia62] Palahenedi H. Diananda. On nonnegative forms in real variables some or all of which are nonnegative. *Mathematical Proceedings of the Cambridge Philosophical Society*, 58:17–25, 1962.

[Dic10] Peter J.C. Dickinson. An improved characterisation of the interior of the completely positive cone. *Electronic Journal of Linear Algebra*, 20:723–729, Nov 2010.

[Dic11] Peter J.C. Dickinson. Geometry of the copositive and completely positive cones. *Journal of Mathematical Analysis and Applications*, 380(1):377–395, 2011.

[Dic13] Peter J.C. Dickinson. On the exhaustivity of simplicial partitioning. *Journal of Global Optimization, in print. DOI: 10.1007/s10898-013-0040-7*, 2013.

[DJL94] John H. Drew, Charles R. Johnson, and Raphael Loewy. Completely positive matrices associated with M-matrices. *Linear and Multilinear Algebra*, 37:303–310, 1994.

[DLC12] Bo Dong, Matthew M. Lin, and Moody T. Chu. Nonnegative rank factorization via rank reduction. *Preprint, submitted. Available at http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.149.5358*, 2012.

[DP13a] Peter J.C. Dickinson and Janez Povh. Moment approximations for set-semidefinite polynomials. *Journal of Optimization Theory and Applications, in print. DOI: 10.1007/s10957-013-0279-7.*, 2013.

[DP13b] Peter J.C. Dickinson and Janez Povh. A new convex reformulation and approximation hierarchy for polynomial optimization. *In construction*, 2013.

[DP13c] Peter J.C. Dickinson and Janez Povh. On a generalization of Pólya's and Putinar-Vasilescu's positivstellensätze. *Preprint, submitted*, 2013.

[DS08] Mirjam Dür and Georg Still. Interior points of the completely positive cone. *Electronic Journal of Linear Algebra*, 17:48–53, 2008.

[Dür10] Mirjam Dür. *Copositive Programming - a Survey,* In: Recent Advances in Optimization and its Applications in Engineering, Moritz Diehl, Francois Glineur, Elias Jarlebring and Wim Michiels (Eds.), pages 3–20. Springer Berlin Heidelberg, 2010.

[EJ08] Gabriele Eichfelder and Johannes Jahn. Set-semidefinite optimization. *Journal of Convex Analysis*, 15:767–801, 2008.

[EJ10] Gabriele Eichfelder and Johannes Jahn. *Foundations of Set-Semidenite Optimization,* In: Nonlinear Analysis and Variational Problems, Panos M. Pardalos, Themistocles M. Rassias and Akhtar A. Khan (Eds.), volume 35 of *Springer Optimization and Its Applications*, pages 259–284. Springer New York, 2010.

[EP12] Gabriele Eichfelder and Janez Povh. On the set-semidefinite representation of nonconvex quadratic programs over arbitrary feasible sets. *Optimization Letters, in print. DOI: 10.1007/s11590-012-0450-3*, 2012.

[GJ79] Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of $\mathcal{NP}$-Completeness.* W. H. Freeman & Co., 1979.

[GL07] Nebojša Gvozdenović and Monique Laurent. Semidefinite bounds for the stability number of a graph via sums of squares of polynomials. *Mathematical Programming*, 110:145–173, 2007.

[GLS88] Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization.* Springer-Verlag (Berlin), 1988.

[GS11] M. Seetharama Gowda and Roman Sznajder. On the non-homogeneity of completely positive cones. *Preprint, submitted. Available at http://www.optimization-online.org/DB_HTML/2012/05/3464.html*, 2011.

[GW80] Leonard J. Gray and David G. Wilson. Nonnegative factorization of positive semidefinite nonnegative matrices. *Linear Algebra and its Applications*, 31:119–127, 1980.

[Hil88] David Hilbert. Über die darstellung definiter formen als summe von formenquadraten. *Mathematische Annalen*, 32(3):342–350, 1888.

[Hil12] Roland Hildebrand. The extreme rays of the 5×5 copositive cone. *Linear Algebra and its Applications*, 437(7):1538–1547, 2012.

[HJ62] Marshall Hall Jr. *Discrete Problems,* In: Survey of Numerical Analysis, John Todd (Ed.), pages 518–542. McGraw-Hill Book Company, Inc., 1962.

[HJR05] Leslie Hogben, Charles R. Johnson, and Robert Reams. The copositive completion problem. *Linear Algebra and its Applications*, 408:207–211, Oct 2005.

[HLP88] Godfrey Harold Hardy, John Edensor Littlewood, and George Pólya. *Inequalities.* Cambridge University Press, 1988.

[HN63] Marshall Hall Jr and Morris Newman. Copositive and completely positive quadratic forms. *Mathematical Proceedings of the Cambridge Philosophical Society*, 59:329–339, 1963.

[Hor76] Reiner Horst. An algorithm for nonconvex programming problems. *Mathematical Programming*, 10(1):312–321, 1976.

[Hor97] Reiner Horst. On generalized bisection of n-simplices. *Mathematics of Computation*, 66(218):691–698, 1997.

[HP73] Alan J. Hoffman and Francisco Pereira. On copositive matrices with $-1, 0, 1$ entries. *Journal of Combinatorial Theory*, 14(3):302–309, May 1973.

[HPT00] Reiner Horst, Panos M. Pardalos, and Nguyen Van Thoai. *Introduction to Global Optimization.* Springer, Dec 2000.

[HUS10] Jean-Baptiste Hiriart-Urruty and Alberto Seeger. A variational approach to copositive matrices. *SIAM Review*, 52(4):593–629, 2010.

[Ikr02] Khakim D. Ikramov. Linear-time algorithm for verifying the copositivity of an acyclic matrix. *Computational Mathematics and Mathematical Physics*, 42:1701–1703, 2002.

[Kar72] Richard M. Karp. *Reducibility among combinatorial problems,* In: Complexity of computer computations, Raymond E. Miller and James W. Thatcher (Eds.), pages 85–103. Plenum Press, New York, 1972.

[Kay87] Mohammad Kaykobad. On nonnegative factorization of matrices. *Linear Algebra and its Applications*, 96:27–33, 1987.

[KB92] Natalia Kogan and Abraham Berman. Characterization of completely positive graphs. *Discrete Mathematics*, 114:297–304, 1992.

[Kea78] Baker Kearfott. A proof of convergence and an error bound for the method of bisection in $\mathbb{R}^n$. *Mathematics of Computation*, 32(144):1147–1153, Oct 1978.

[KG12] Vassilis Kalofolias and Efstratios Gallopoulos. Computing symmetric nonnegative rank factorizations. *Linear Algebra and its Applications*, 436:421–435, 2012.

[KM40] Mark Krein and David Milman. On extreme points of regular convex sets. *Studia Mathematica*, 9:133–138, 1940.

[Koe98] Wolfram Koepf. *Hypergeometric Summation: An Algorithmic Approach to Summation and Special Function Identities.* Friedrich Vieweg & Sohn Verlagsgesellschaft mbH, 1998.

[KP02] Etienne de Klerk and Dmitrii V. Pasechnik. Approximation of the stability number of a graph via copositive programming. *SIAM Journal on Optimization*, 12(4):875–892, 2002.

[Las10] Jean-Bernard Lasserre. *Moments, Positive Polynomials and Their Applications.* World Scientific, 2010.

[Las11] Jean-Bernard Lasserre. A new look at nonnegativity on closed sets and polynomial optimization. *SIAM Journal on Optimization*, 21(3):864–885, 2011.

[Las13] Jean-Bernard Lasserre. New approximations for the cone of copositive matrices and its dual. *Mathematical Programming, in print. DOI: 10.1007/s10107-013-0632-5*, 2013.

[Lau09] Monique Laurent. *Sums of Squares, Moment Matrices and Optimization Over Polynomials,* In: Emerging Applications of Algebraic Geometry, Mihai Putinar and Seth Sullivant (Eds.), volume 149 of *The IMA Volumes in Mathematics and its Applications*, pages 157–270. Springer New York, 2009.

[LKF04] Yingbo Li, Anton Kummert, and Andreas Frommer. A linear programming based analysis of the CP-rank of completely positive matrices. *International Journal of Applied Mathematics and Computer Science*, 14(1):25–31, 2004.

[LVBL98] Miguel Sousa Lobo, Lieven Vandenberghe, Stephen Boyd, and Hervé Lebret. Applications of second-order cone programming. *Linear Algebra and its Applications*, 284(1–3):193–228, 1998.

[Mar71] Thomas L. Markham. Factorizations of completely positive matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 69:53–58, 1971.

[MK87] Katta G. Murty and Santosh N. Kabadi. Some $\mathcal{NP}$-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39(2):117–129, 1987.

[MM62] John E. Maxfield and Henryk Minc. On the matrix equation $X'X = A$. *Proceedings of the Edinburgh Mathematical Society*, 13(2):125–129, 1962.

[Mot52] Theodore S. Motzkin. Copositive quadratic forms. *National Bureau of Standards Report*, 1818:11–22, 1952.

[MS65] Theodore S. Motzkin and Ernst G. Straus. Maxima for graphs and a new proof of a theorem of Turán. *Canadian Journal of Mathematics*, 17:533–540, 1965.

[Mur03] Kazuo Murota. *Discrete Convex Analysis.* SIAM, 2003.

[NN93] Yurii Nesterov and Arkadi Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming.* SIAM, 1993.

[NTZ11] Karthik Natarajan, Chung Piaw Teo, and Zhichao Zheng. Mixed 0-1 linear programs under objective uncertainty: A completely positive representation. *Operations Research*, 59(3):713–728, 2011.

[Par00] Pablo A Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization.* PhD thesis, California Institute of Technology, 2000.

[PV99] Mihai Putinar and Florian-Horia Vasilescu. Positive polynomials on semialgebraic sets. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, 328(7):585–589, 1999.

[PVZ07] Javier F. Peña, Juan C. Vera, and Luis F. Zuluaga. Computing the stability number of a graph via linear and semidefinite programming. *SIAM Journal on Optimization*, 18(1):87–105, 2007.

[QKRT98] Arie J. Quist, Etienne de Klerk, Cornelis Roos, and Tamàs Terlaky. Copositive relaxation for general quadratic programming. *Optimization Methods & Software*, 9:185–208, 1998.

[Rob73] Raphael M. Robinson. *Some definite polynomials which are not sums of squares of real polynomials,* In: Selected questions of algebra and logic (collection dedicated to the memory of A. I. Mal'cev) (Russian), Yu. L. Ershov et al. (Eds.), pages 264–282. Izdat. 'Nauka' Sibirsk. Otdel., Novosibirsk, 1973.

[Roc70] R. Tyrrell Rockafellar. *Convex Analysis.* Princeton University Press, 1970.

[Roy63] Halsey L. Royden. *Real Analysis.* The Macmillan Company, New York, 2nd edition, 1963.

[RTV97] Cornelis Roos, Tamás Terlaky, and Jean-Philippe Vial. *Theory and algorithms for linear optimization: An Interior Point Approach.* Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Ltd., 1997.

[SM09] Naomi Shaked-Monderer. A note on upper bounds of the cp-rank. *Linear Algebra and its Applications*, 431:2407–2413, 2009.

[SMBJS12] Naomi Shaked-Monderer, Immanuel M. Bomze, Florian Jarre, and Werner Schachinger. On the cp-rank and the minimal cp factorizations of a completely positive matrix. *Preprint, available at http://www.optimization-online.org/DB_HTML/2012/06/3517.html*, 2012.

[SS00] Alexander Shapiro and Katya Scheinberg. *Duality and Optimality Conditions,* In: Handbook of Semidefinite Programming, Henry Wolkowicz, Romesh Saigal and Lieven Vandenberghe (Eds.), volume 27, pages 67–110. Springer US, 2000.

[TH88] Hoang Tuy and Reiner Horst. Convergence and restart in branch-and-bound algorithms for global optimization. Application to concave minimization and D.C. optimization problems. *Mathematical Programming*, 41(1):161–183, 1988.

[Tuy91a] Hoang Tuy. Effect of the subdivision strategy on convergence and efficiency of some global optimization algorithms. *Journal of Global Optimization*, 1(1):23–36, 1991.

[Tuy91b] Hoang Tuy. Normal conical algorithm for concave minimization over polytopes. *Mathematical Programming*, 51(1):229–245, 1991.

[WSV00] Henry Wolkowicz, Romesh Saigal, and Lieven Vandenberghe. *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications.* Springer, 2000.

[XL00] Changqing Xu and Jiong-seng Li. A note on completely positive graphs. *Systems Science and Mathematical Sciences*, 13:121–125, 2000.

[ZL00] Xiao-dong Zhang and Jiong-seng Li. Completely positive matrices having cyclic graphs. *Journal of Mathematical Research & Exposition*, 1:27–31, 2000.

Nomenclature

1. Optimisation

Val(P)	Optimal value for an optimisation problem (P), page 10.
Feas(P)	Set of feasible points for an optimisation problem (P), page 10.
Opt(P)	Set of optimal solutions for an optimisation problem (P), page 10.
s.t.	such that / subject to.

2. Vectors

$\mathbb{R}^n$	Set of real n-vectors.
$\mathbb{R}^n_+$	Set of nonnegative real n-vectors, also referred to as the *Nonnegative orthant*.
$\mathbb{R}^n_{++}$	Set of strictly positive real n-vectors.
$\mathbb{Q}^n$	Set of rational n-vectors. (Similarly to $\mathbb{R}^n$ for $\mathbb{Q}^n_+$ and $\mathbb{Q}^n_{++}$.)
$\mathbb{Z}^n$	Set of integer n-vectors. (Similarly to $\mathbb{R}^n$ for $\mathbb{Z}^n_+$ and $\mathbb{Z}^n_{++}$.)
$\mathbb{N}^n_{\leq d}$	This is defined as $\mathbb{N}^n_{\leq d} := \{\boldsymbol{\alpha} \in \mathbb{Z}^n_+ \mid \boldsymbol{e}^\mathsf{T}\boldsymbol{\alpha} \leq d\}$.
$\mathbb{N}^n_{=d}$	This is defined as $\mathbb{N}^n_{=d} := \{\boldsymbol{\alpha} \in \mathbb{Z}^n_+ \mid \boldsymbol{e}^\mathsf{T}\boldsymbol{\alpha} = d\}$.

For vector sets above, the "n" is excluded if the dimension is equal to 1.

$\mathbb{R}^{\mathcal{A}}$	For a set $\mathcal{A}$, this is the set of real vectors of order $\lvert\mathcal{A}\rvert$, indexed by elements in $\mathcal{A}$.
$\boldsymbol{e}$	Vector of all-ones.
$\boldsymbol{e}_i$	Vector with ith entry equal to 1 and all other entries equal to 0.
$\langle \mathbf{a}, \mathbf{b} \rangle$	For two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, we define their inner product as $\langle \mathbf{a}, \mathbf{b} \rangle := \mathbf{a}^\mathsf{T}\mathbf{b} = \sum_i (\mathbf{a})_i (\mathbf{b})_i$.
$\lVert\mathbf{a}\rVert_2$	For $\mathbf{a} \in \mathbb{R}^n$, define its 2-norm as $\lVert\mathbf{a}\rVert_2 := \sqrt{\langle \mathbf{a}, \mathbf{a} \rangle} = \sqrt{\sum_i (\mathbf{a})_i^2}$.

3. Matrices

$\mathcal{S}^n$	Set of symmetric matrices of order n.
$\mathcal{C}^n$	Set of copositive matrices of order n, page 3.
$\mathcal{C}^{*n}$	Set of completely positive matrices of order n, page 3.
$\mathcal{S}_+^n$	Set symmetric positive semidefinite matrices of order n, page 4.
$\mathcal{N}^n$	Set symmetric nonnegative matrices of order n.
$\mathcal{S}_+^n \cap \mathcal{N}^n$	Set of doubly nonnegative matrices of order n, page 4.

For matrix sets above, the "n" is excluded if order is apparent from the context.

E	Matrix of all-ones.
I	Identity matrix.
E_{ij}	This is equal to $\boldsymbol{e}_i\boldsymbol{e}_i^\mathsf{T}$ for $i = j$ and $(\boldsymbol{e}_i\boldsymbol{e}_j^\mathsf{T} + \boldsymbol{e}_j\boldsymbol{e}_i^\mathsf{T})$ for $i \neq j$.
$\mathrm{Diag}(\mathbf{d})$	For $\mathbf{d} \in \mathbb{R}^n$, this gives $D \in \mathcal{S}^n$ such that $(D)_{ii} = (\mathbf{d})_i$ for all i and $(D)_{ij} = 0$ for all $i \neq j$.
$\mathrm{diag}(A)$	For $A \in \mathbb{R}^{n\times n}$, this gives $\mathbf{a} \in \mathbb{R}^n$ such that $(\mathbf{a})_i = (A)_{ii}$ $\forall i$.
$\mathrm{rank}(A)$	Rank of a matrix A.
$\mathrm{trace}(A)$	Trace of a matrix A.
$\det(A)$	Determinant of a matrix A.
$\mathrm{Ker}(A)$	Kernel of a matrix A.
$\text{cp-rank}(A)$	The cp-rank of a matrix A, page 49.
$\mathcal{V}^A$	Set of Zeros for $\mathbf{x}^\mathsf{T}A\mathbf{x}$ in the Nonnegative Orthant, page 77.
$S(\boldsymbol{\theta})$	Hildebrand Matrices, page 91.
$\langle A, B\rangle$	For two matrices $A, B \in \mathbb{R}^{n\times m}$, we define their inner product as $\langle A, B\rangle := \mathrm{trace}(A^\mathsf{T}B) = \sum_{i,j}(A)_{ij}(B)_{ij}$.
$\Vert A\Vert_2$	For $A \in \mathbb{R}^{n\times m}$, define its 2-norm as $\Vert A\Vert_2 := \sqrt{\langle A, A\rangle} = \sqrt{\sum_{i,j}(A)_{ij}^2}$.

4. Sets

$\subseteq$	For sets $\mathcal{A}, \mathcal{B}$, we have $\mathcal{A} \subseteq \mathcal{B}$ if and only if there does not exist $\mathbf{x} \in \mathcal{A}$ such that $\mathbf{x} \notin \mathcal{B}$.
$\subset$	For sets $\mathcal{A}, \mathcal{B}$, we have $\mathcal{A} \subset \mathcal{B}$ if and only if $\mathcal{A} \subseteq \mathcal{B}$ and $\mathcal{A} \neq \mathcal{B}$.
$\mathrm{span}(\mathcal{M})$	Span of a set $\mathcal{M}$, page 7.
$\mathrm{aff}(\mathcal{M})$	Affine hull of a set $\mathcal{M}$, page 7.

$\operatorname{conv}(\mathcal{M})$	Convex hull of a set $\mathcal{M}$, page 7.
$\operatorname{cone}(\mathcal{M})$	For a set $\mathcal{M}$, define $\operatorname{cone}(\mathcal{M}) := \{\alpha\mathbf{x} \mid \alpha \in \mathbb{R}_+,\ \mathbf{x} \in \mathcal{M}\}$.
$\operatorname{conic}(\mathcal{M})$	Conic hull of a set $\mathcal{M}$, page 7.
$\operatorname{cl}(\mathcal{M})$	Closure of a set $\mathcal{M}$, page 7.
$\operatorname{int}(\mathcal{M})$	Interior of a set $\mathcal{M}$, page 7.
$\operatorname{bd}(\mathcal{M})$	Boundary of a set $\mathcal{M}$, page 7.
$\operatorname{reint}(\mathcal{M})$	Relative interior of a set $\mathcal{M}$, page 7.
$\operatorname{rbd}(\mathcal{M})$	Relative boundary of a set $\mathcal{M}$, page 7.
$\operatorname{recc}(\mathcal{M})$	Recession cone of a set $\mathcal{M}$, page 7.
$\dim(\mathcal{M})$	Dimension of a set $\mathcal{M}$, page 7.
$\mathcal{M}^*$	Dual of a set $\mathcal{M}$, page 12.
$\mathrm{S}(\mathcal{M},\varepsilon)$	For $\varepsilon \in \mathbb{R}_{++}$, this is the ε outer approximation of $\mathcal{M}$, page 36.
$\mathrm{S}(\mathcal{M},-\varepsilon)$	For $\varepsilon \in \mathbb{R}_{++}$, this is the ε inner approximation of $\mathcal{M}$, page 36.
$\operatorname{Ext}(\mathcal{K})$	Generators of extreme rays of a proper cone $\mathcal{K}$, page 87.
$\operatorname{Exp}(\mathcal{K})$	Generators of exposed rays of a proper cone $\mathcal{K}$, page 87.
$\mathcal{F}(\mathcal{K},\mathbf{v})$	Exposed face of a proper cone $\mathcal{K}$ given by the intersection with hyperplane through the origin whose normal is $\mathbf{v} \in \mathcal{K}^*$, page 87.

5. Graphs

$\alpha(G)$	Stability number of a simple graph G, page 21.
A_G	Adjacency matrix of a simple graph G, page 22.
$G(A)$	Underlying graph of the symmetric matrix A, page 51.

6. Functions and Moments

$\mathbf{x}^{\boldsymbol{\alpha}}$	For $\mathbf{x} \in \mathbb{R}^n$, $\boldsymbol{\alpha} \in \mathbb{Z}^n_+$, we have $\mathbf{x}^{\boldsymbol{\alpha}} := \prod_{i=1}^n x_i^{\alpha_i}$ (where $0^0 := 1$).
$\mathbf{v}_d(\mathbf{x})$	$\mathbf{v}_d$ is function from $\mathbb{R}^n$ to $\mathbb{R}^{\mathbb{N}^n_{\le d}}$ such that $\mathbf{v}_d(\mathbf{x}) := (\mathbf{x}^{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n_{\le d}}$ for all $\mathbf{x} \in \mathbb{R}^n$.
$\mathbf{u}_d(\mathbf{x})$	$\mathbf{u}_d$ is function from $\mathbb{R}^n$ to $\mathbb{R}^{\mathbb{N}^n_{=d}}$ such that $\mathbf{u}_d(\mathbf{x}) := (\mathbf{x}^{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n_{=d}}$ for all $\mathbf{x} \in \mathbb{R}^n$.
$\mathbb{R}[\mathbf{x}]$	The ring of polynomials in the variables $\mathbf{x} = (x_1,\dots,x_n)$ with real coefficients, page 125.
$\deg(f)$	Degree of a polynomial f, page 125.

$f^{-1}(\mathcal{A})$	For a function f from $\mathbb{R}^n$ to $\mathbb{R}$ and a set $\mathcal{A} \subseteq \mathbb{R}$, we define $f^{-1}(\mathcal{A}) := \{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) \in \mathcal{A})$.
$f^{-1}(\alpha)$	For a function f from $\mathbb{R}^n$ to $\mathbb{R}$ and a scalar $\alpha \in \mathbb{R}$, we define $f^{-1}(\alpha) := f^{-1}\big(\{\alpha\}\big) = \{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) = \alpha\}$.
$\mathrm{support}(\mu)$	Support of a Borel measure μ, page 126.
$y^{\mu}_{\boldsymbol{\alpha}}$	$\boldsymbol{\alpha}$ moment of a Borel measure μ, page 126.
$\mathbf{y}^{\mu}$	Infinite dimensional vector $(y^{\mu}_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{Z}^n_+}$.
$\mathcal{M}_{\mathrm{B}}$	Set of Borel measures such that their moments exist, page 126.
$M(f\mathbf{y})$	Localising matrix associated with a function f and an infinite dimensional vector $\mathbf{y}$, page 127.
$M_d(f\mathbf{y})$	Localising matrix of order d, pages 127 and 135.
$M_{=d}(f\mathbf{y})$	Certain submatrix of localising matrix, page 135.
$\mathcal{M}_{\leq d}$	This is defined as the set $\left\{\mathbf{y} \in \mathbb{R}^{\lvert \mathbb{N}^n_{\leq 2d} \rvert} \,\middle|\, M_d(\mathbf{y}) \in \mathcal{S}_+\right\}$, page 135.
$\mathcal{M}_{=d}$	This is defined as the set $\left\{\mathbf{y} \in \mathbb{R}^{\lvert \mathbb{N}^n_{=2d} \rvert} \mid M_{=d}(\mathbf{y}) \in \mathcal{S}_+\right\}$, page 135.
SOS	Sum-of-squares, page 135.
$\Sigma_{\leq d}$	This is defined as the set $\left\{\mathbf{f} \in \mathbb{R}^{\lvert \mathbb{N}^n_{\leq 2d} \rvert} \,\middle|\, f \text{ is SOS}\right\}$, page 135.
$\Sigma_{=d}$	This is defined as the set $\left\{\mathbf{f} \in \mathbb{R}^{\lvert \mathbb{N}^n_{=2d} \rvert} \mid f \text{ is SOS}\right\}$, page 135.
$\mathcal{K}^r_n$	The Parrilo-r cone, page 137.
$\mathcal{Q}^r_n$	An approximation hierarchy for the copositive cone, page 145.
$\mathcal{C}^r_n$	An approximation hierarchy for the copositive cone, page 146.

Index

2-norm, 7

Acyclic matrix, *see* Graphs of a matrix
Adjacency matrix, *see* Graph
Affine hull/set, 8
Approximation
  Convergent, 105
  $\mathcal{C}_n^r$, 146
  Heirarchy, 105
  $\mathcal{K}_n^r$, *see* Sum-of-squares
  Moment, *see* Moment
  Optimisation, 106
  Parrilo cone, *see* Sum-of-squares
  Polynomial Optimisation, *see* Polynomial
  $\mathcal{Q}_n^r$, 145
  Simplex, *see* Simplex

Base, 15
Boundary, 8
  Relative boundary, 8
Bounded set, 9

Carathéodory's theorem, 10
Chains, *see* Graphs of a matrix
Circular matrix, *see* Graphs of a matrix
Clique, *see* Graph
Closed set, 8
Closure, 8
co-$\mathcal{NP}$, *see* Complexity
Completely positive, 3
  Completely positive graph, 53
  Complexity, *see* Complexity
  Copositive optimisation, 17
  Dual, 15
  Exposed ray, 90
  Extreme ray, 90
  Facet, 99
  Interior, 79–84
  Maximal face, 98–102
  Order 2, 90
  Order 5, 101
  Proper cone, 75–76
Complexity, 35–47
  co-$\mathcal{NP}$, 36, 37
  $\mathcal{C}^*$ membership, 41
  Copositive membership, 41
  Copositive optimisation, 21
  Ellipsoid method, 37–41
  ε inner/outer approximation, 36
  $\mathcal{NP}$, 36, 37
  $\mathcal{NP}$-hard, 37
  Quintuple, 37
  Strong membership (MEM), 35
  Strong separation, 35
  Weak membership (WMEM), 36
Component submatrix, *see* Graphs of a matrix
Cone, 8
Conic hull, 8
Conic optimisation, *see* Optimisation
Connected matrix, *see* Graphs of a matrix
Convex hull/set, 8
Copositive, 3
  Complexity, *see* Complexity
  Copositive optimisation, 17
  Dual, 15
  Exposed ray, 93
  Extreme ray, 91–94
  Facet, 97
  Interior, 79
  Maximal face, 95–97
  Order 2, 90
  Order 5, 138–139
  Proper cone, 75–76
cp-rank, *see* Rank-one decomposition set

$\mathcal{C}_n^r$, 146

Degree of an index, *see* Graphs of a matrix
Dimension, 8, 97, 98, 100
Doubly nonnegative, 4
  Dual, 15
Dual, 12–16
Duality, *see* Optimisation

Ellipsoid method, *see* Complexity
ε inner/outer approximation, *see* Complexity
Exposed point, 85
Extreme point, 85

Face, 85
  Exposed face, 85, 87
  Facet, 85
    Completely positive, 99
    Copositive, 97
  Maximal face, 86, 88
    Completely positive, 98–102
    Copositive, 95–97
Facet, *see* Face
Feasible set, *see* Optimisation
Full-dimensional set, 8

Generator, *see* Ray
Geometry, 7–10, 85–89
Graph
  Adjacency matrix, 22
  Clique, 18
    Maximum weight clique, 18
  Stable set, 21
    Stability number, 21
Graphs of a matrix, 51
  Acyclic matrix, 51, 57, 71
  Chains, 57–59, 64–68
  Circular matrix, 51, 60–63, 71
  Component submatrix, 51
  Connected matrix, 51
  Degree of an index, 51
  Underlying graph of A, 51
  Weighted-graph of A, 51

Hildebrand matrices, 91, 139
Horn matrix, 92, 98

Inner product, 7
Interior, 8
  Completely positive, 79–84
  Copositive, 79
  Relative interior, 8

$\mathcal{K}_n^r$, *see* Sum-of-squares

Linear space, 7

Membership, *see* Complexity
Moment, 126–127
  Copositive, 131–133
  Dual, 135–136
  Localising matrix, 127
  Measure
    Borel measure, 126
    Finite Borel measure, 127
    Probability measure, 126
    Support, 126
  Set-semidefinite, 128, 130, 131

$\mathcal{NP}$, *see* Complexity

Optimisation, 10–18
  Attained, 11
  Conic optimisation, 11
    Copositive optimisation, 17
    Duality, 12–16
    Linear optimisation, 17
    Semidefinite optimisation, 17
  Feasible set, 11
  Optimal solutions, 11
  Optimal value, 11
  Programming, 12
  Quadratic binary optimisation, 27
  Slater's condition, 13
  Standard quadratic optimisation, 26

Parrilo cone, *see* Sum-of-squares
Pointed set, 9
Polynomial, 125, 147
  Constraints, 147
  Optimisation, 147–149
    Approximation, 149
    Convex, 148
Positive semidefinite, 4
  Dual, 15

Semidefinite optimisation, 17
Positivstellensatz, 145–146
Principal submatrix, 5
Programming, *see* Optimisation
Proper cone, 9
Completely positive, 75–76
Copositive, 75–76

$\mathcal{Q}_n^r$, 145
Quintuple, *see* Complexity

Rank-one decomposition set, 49
cp-rank, 49
Minimal, 50, 69–71
Ray, 86
Exposed ray, 87, 88
Completely positive, 90
Copositive, 93
Extreme ray, 87, 88
Completely positive, 90
Copositive, 91–94
Generator, 86
Recession cone, 8

Separation, *see* Complexity
Set of zeros, 77–78, 100
Set-semidefinite, 23–33, 125–126
Quadratic binary optimisation, 27
Simplex
Copositivity, 108–110, 122
Diameter, 107
Edges, 107
Partition, 108
Diameter, 108
Exhaustive, 108, 118–120
Nested, 108
Partitioning methods, 110–113
Bisection, 110, 111
Controlling bisection, 112
Free bisection, 112
Free partition, 113
ω-subdivisions, 112
Radial subdivision, 112
Standard simplex, 108
Triangle, 107
Vertices, 107
Slater's condition, *see* Optimisation
Span, 7
Stability number, *see* Graph
Stable set, *see* Graph
Straszewicz's theorem, 87
Strong membership (MEM), *see* Complexity
Strong Separation (SEP), *see* Complexity
Sum-of-squares, 135
Dual, 135–136
Parrilo cone, $\mathcal{K}_n^r$, 137
Convergence, 146
Dual, 141–143
Scaling, 137–141

Underlying graph of A, *see* Graphs of a matrix

Weak membership (WMEM), *see* Complexity

Zeros, *see* Set of Zeros